

Mani Sarkar

github: [neomatrix369](https://github.com/neomatrix369)

twitter: [@theNeomatrix369](https://twitter.com/theNeomatrix369)

blogs: <https://medium.com/@neomatrix369>



From backend development to machine learning

my journey from Software Engineering to ML/Data Science

3rd July 2020

hashtag: **#abhishektalks**

About me



Mani Sarkar

[More about me](#)

**Freelance Software,
Data, ML Engineer**

Java / JVM

**Cloud / Infra /
DevOps**

Polyglot developer

**LJC, Devoxx,
developer communities**

**Code quality, testing,
performance, DevOps, deep
affinity for AI/ML/DL, NN...**

***Strengthening* teams and
helping them *accelerate***

**JCP member, F/OSS projects:
@adoptopenjdk @graalvm
@truffleruby**

**Java Champion, Oracle Groundbreaker Ambassador,
Software Crafter, Blogger, Speaker**

@theNeomatrix369

I'm a big fan of...

Nikola Tesla and his fine work

Richard Feynman, his books and quotes!

Sir Arthur Conan Doyle and his work (I have all his 65+ series on DVD)

It's important
to have heros
and mentors!

Agenda

About the talk

- *Timeline: Backend dev to ML Engineer*
- *Demos*
- *References to other talks*
- *Resources*
- *Summary and Closing*
- *Q&A*
- *Appendix section: more good stuff for later*



Presentation slides: *live*

<https://bit.ly/backend-dev-to-ml-slides>



Thank You!

- **Abhishek Thakur**, for organising this session, and giving me a chance to present at this forum
- And to “**you**”, for sparing your valuable time and trusting me

So honoured!



Disclaimer

- *YMMV*
- Might have rough edges and **inaccuracies**
- Sharing our **learnings** over the past years
- Gathered ideas from **different sources**
- **Sharing ideas and experiences**

Citation

The respective authors and creators are, and remain the true owners of the images and other artifacts used in this presentation.

Thank you for your creations!

Introduction

Summarising
~20 years till 2017

Dos

First PC: 8086/88

QuickBasic

VBA macros

Games: PacMan, Digger,
WITWI Carmen
San Diego, Prince of
Persia, Myst, among
others

Object Pascal / Delphi

SmallTalk

Windows

Scala

R

Java

Linux

Bash shell

macOS

During the first half of my
career I have been writing
Desktop, command-line,
REST-enabled and
server-side apps on the
Windows platform, and
later doing the same on
Linux and macOS.



@theNeomatrix369

Skipping a lot of details,
lots to talk about those
~20 years of my career



~2009 / ~2010:

Throwing myself into the
deep-end of tech!

**Stopped using Windows
(secondary OS on the partition)**

**Very serious meetup go-er - sometimes
*4 times a week (sometimes weekends)***

**Throw myself in the deep-end: Linux
CLI and macOS**

Introduction to Git

**Reading a
lot**

Getting more serious about Java

**Interest and learning about VMs,
Interpreters, Compilers, chipsets,
GPUs**

**Building OpenJDK (using make and
bash)**

Running VMs: VirtualPC, vmWare

@theNeomatrix369

~2013 - ~2017:

A new phase begins

**OpenJDK =>
AdoptOpenJDK**

Core Java

Scala

JVM internals

Performance Tuning / GC

Shell scripting

**Helping
Conference
organisers**

**Workshops and
speaking at
conferences**

VBA Macros

Writing blogs

Wrote R scripts for ~6 months

My first R blog post (~2013)

Developing interest in data and ML

@theNeomatrix369

**Agile
methodology**

Code quality

Tests and TDD

Code coverage

**Craftmanship
(crafter)**

Fuzzy testing

Refactoring code

Mutation testing

**Property-based
testing**

I was learning what it
takes to become a
better professional!

- **Speak online or at conferences**
- **Write blogs or articles or papers**
- **Push code to git (personal or F/OSS projects)**
- **Review code, books, articles, etc...**
- **....just help others become better at what they do**
- **Collaborate with other developers**
- **Be organised, practical and develop pragmatism**
- **Read and write a lot of code**
- **Rewrite any or all of the above (build muscle memory)**
- **and keep doing them iteratively, and more...**

@theNeomatrix369



~2017 / 2018: another
important phase

Switched to Freelancing

Best Practices

Moved roles

**Small/Startup clients
only**

Small teams only

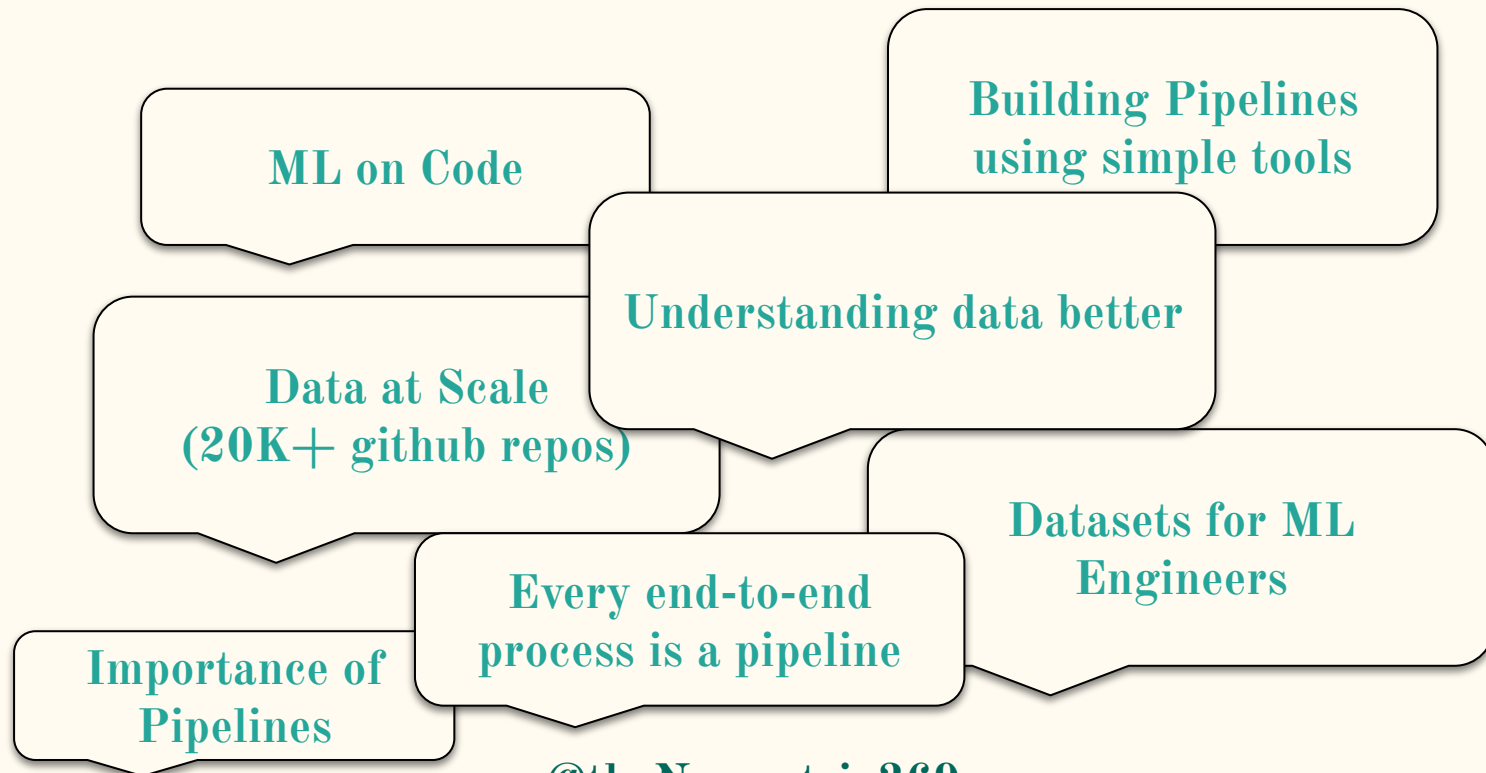
**Accelerate development
process**

MVPs and PoCs

**Development
methodologies**

@theNeomatrix369

Working on a ML/Data Engineering project



@theNeomatrix369

AI/ML/DL: Learn and collaborate

Joining Meetups

Attending
Conferences

Participating in Study
groups / AI Labs

Focus moved to more
AI/ML/DL topics

WhatsApp, Telegram,
Slack channels, Mailing
list, etc...

Bravo!



Gave talks at UCL's CREST Open Workshops

See awesome-graal work

My talks at the
CREST workshops:
[1] [2]

Discussing and developing
ML on Code ideas

~2018 till date:
AI/ML/DL fun begins

Thanks to Yolande Poirier for
posting all those AI/ML/DL
links on @java between 2016
and 2018

AI/ML/DL git repo

@theNeomatrix369

<https://github.com/neomatrix369/awesome-ai-ml-dl>

README.md

Awesome AI-ML-DL

awesome License CC BY-SA 4.0

Better NLP: docker pulls 55

NLP Java: docker pulls 310 | NLP Clojure: docker pulls 8 | NLP Kotlin: docker pulls 9 | NLP Scala: docker pulls 6 |

NLP using DL4J (cuda) docker pulls 88

Dataiku DSS: docker pulls 26 | Grakn: docker pulls 1.7k | Jupyter-docker pulls 25 |

MLPMNist using DL4J: docker pulls 46 | Zeppelin: docker pulls 62

No packages published
Publish your first package

Contributors 15

+ 4 contributors

Please watch
repo to follow
latest updates
and commits!

Please ***star***, **watch**,
fork and share the
repo, as it motivates
me to continue with
my work

@theNeomatrix369

AI/ML/DL artefacts

@theNeomatrix369

[Cheatsheets](#)

[Notebooks](#)

[Presentations](#)

[Examples](#)

[Virgilio](#)

[Guides](#)

[Docker
containers](#)

[Study Notes](#)

[Things to know list](#)

@theNeomatrix369

Thanks to the Meet-a-Mentor
initiative by LJC during
2018-19, for holding the ML
Study Group!

Talk on data

During this time I attended the study group
I paired with Jeremie Charlet to write a talk
on data, see Do we know our data, as well as
we know our tools? talk!

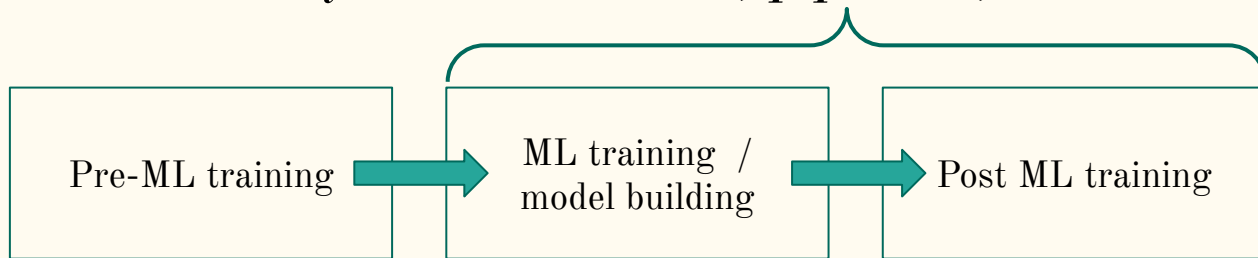
Yes, yes I was there!
Recommend it.



Developing ideas about data and ML

Fan of
AutoML!

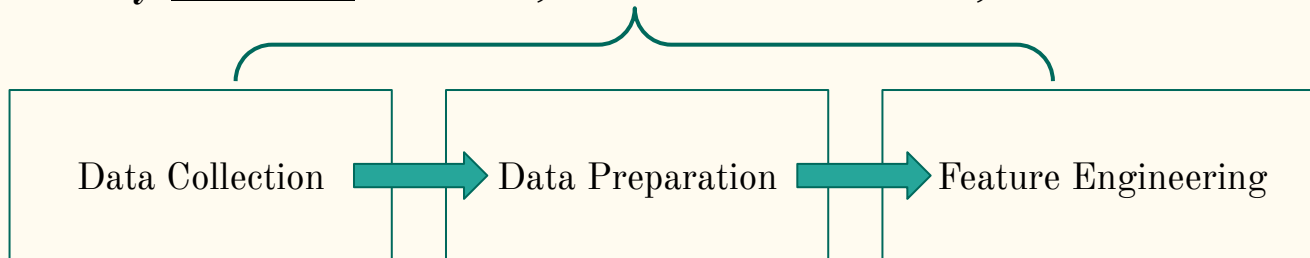
Mostly solved: AutoML, pipelines, etc...



Developing ideas about data and ML

So much
to
achieve...

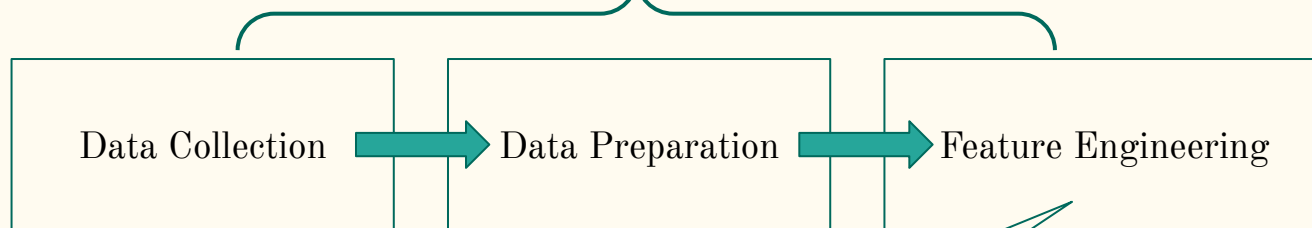
Mostly unsolved: manual, human intervention, some tools



Developing ideas about data and ML

EDA, due diligence

Very important steps, more important than just building model



When done well gives
reliable and robust
models

Critical
thinking

Thanks to Yaz for the
Tensorflow meetups in
London, UK during 2018 and
2019

@theNeomatrix369

My Study
notes from
Tensorflow
and other
topics

Rewrote a
Keras post:
BoW,
Classification

..○
Please have a look at all the TPU
notebooks on each of the links!

Followed on by a post

Enjoyed CPU
vs GPU vs
TPU
comparisons

on TPUs
○ ○ ○

How to harness the Powers of the Cloud TPU

Thanks to Professor Ajit
Jaokar for his meetups and
the AI Labs in London, UK
during 2018 and 2019

@theNeomatrix369

Meetups and AI Labs

- Machine Learning is Fun! See [Book](#) | [Tutorial / Blogs](#)
- [Better NLP](#) library launch, see presentations:
 - First presentation: launch of Better NLP
 - Follow-up presentation of Better NLP

Should I say the
book is highly
recommended



And many others in my
journey...

@theNeomatrix369

Awesome AI/ML/DL repo

Thanks to the contributors for your
pull requests

Collaboration with
MWML 

Working with
Virgilio to make
better guides

We are
cross-linking
resources with
MWML &
Virgilio



All things AI/ML/DL
I'm excited about...

@theNeomatrix369

NLP

**Metrics: tracking and
analysis**

Bayesian concepts

Maths

ML on Code

**Model and experiment
result analysis**

Visualisation

Feature Engineering

Probability

Data, Numbers and distributions

**Uncertainty
concepts**

**Some of the fundamental ideas behind
modelling perfection**

Notebooks/kernels...

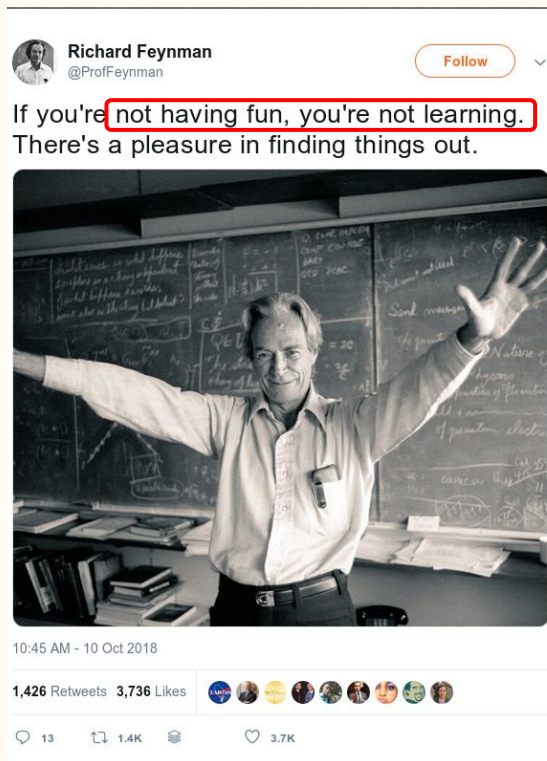
Really enjoyed
doing them!

- A good way to learn anything...
- Here's a list of notebooks/kernels I rewrote/refactored/improved:
 - [Kaggle kernels](#)
 - Google Colab notebooks
 - Tensorflow [GPU & TPU](#) | [TPU](#)
 - Generative Image: [CPU](#) | [GPU](#) | [TPU](#)
 - [Resources on AI/ML/DL](#)



Have fun!

See more Feynman
quotes here



@theNeomatrix369

Participating in Competitions

- Late 2019: participating in competitions
- Become an **Kaggle Discussion Expert** two weeks ago
- Won my first **Bronze medal** at the recent Kaggle competition earlier this week
- See other results from online participations
- **Thanks:** Carlo, Marco, Abhishek and many others for your collaborations and inspirations

It's so much fun!



Demo: RoBERTa

About the demo: RoBERTa

- **Code on GitHub:**

- <https://github.com/neomatrix369/chatbot-conversations>

- Simple chatbot like app written in Python
- Runs as a Flask app (REST endpoint server)
- Demonstrates some NLP capabilities
- Download Roberta models
- Uses RoBERTa Base (others can be used as well)
- We wrote a similar one in [Java using Eliza](#)

```
File Edit View Search Terminal Tabs Help
./run-helidon-world.sh x satyasai@sai-XPS-15-9560: ~/git-repos/ai-ml-di-stuff/ch... x
Other world: were discussing you -- not me.
Helidon: We were discussing you -- not me.
Other world: We were discussing you -- not
Helidon: We were discussing you -- not me.
Other world: We were him -- not me.
Helidon: I rather hear from you first.
Other world: I must hear you first.
Helidon: You say must hear I first?
Other world: You say hear it first
Helidon: We were discussing you -- not me.
Other world: We were discussing you --
Helidon: Oh, I ?
Other world: Oh, and
Helidon: Please go on.
Other world: Please move on.
Helidon: I'm all ears!
Other world: I'm all set
Helidon: Can you elaborate on that?
Other world: Can you on...
Helidon: What are your feelings now?
Other world: What your options now
Helidon: Why do you ask?
Other world: How you ask
Helidon: What comes to mind when you ask that?
Other world: What comes to mind when hearing that
Helidon: Please go on.
Other world: To go on.
Helidon: I rather hear from you first.
Other world: I rather from the first.
Helidon: Do you say rather from the first for some special reason?
```

```
File Edit View Search Terminal Tabs Help
./run-helidon-world.sh x satyasai@sai-XPS-15-9560: ~/git-repos/ai-ml-di-stuff/ch... x
Other world: Why do you ask?
Roberta: How you ask? [confidence: 1.294%]
127.0.0.1 - - [02/Jul/2020 23:32:04] "GET /send?message=Why%20do%20y
Other world: What comes to mind when you ask that?
Roberta: What comes to mind when hearing that? [confidence: 31.04%]
127.0.0.1 - - [02/Jul/2020 23:32:12] "GET /send?message=What%20comes
Other world: Please go on.
Roberta: To go on. [confidence: 13.759%]
127.0.0.1 - - [02/Jul/2020 23:32:20] "GET /send?message=Please%20go%
Other world: I rather hear from you first.
Roberta: I rather from the first. [confidence: 63.393%]
127.0.0.1 - - [02/Jul/2020 23:32:28] "GET /send?message=I%20rather%2
Other world: Do you say rather from the first for some special reason?
Roberta: And you say rather from the first for some reason? [confidence: 5.095%]
127.0.0.1 - - [02/Jul/2020 23:32:36] "GET /send?message=Do%20you%20s
reason? HTTP/1.1" 200 -
Other world: We were discussing you -- not me.
Roberta: We were - you -- me. [confidence: 5.095%]
127.0.0.1 - - [02/Jul/2020 23:32:44] "GET /send?message=We%20were%20
```

```
file Edit View Search Terminal Tabs Help
./run-helidon-world.sh x satyasai@sai-XPS-15-9560: ~/git-repos/ai-ml-di-stuff/ch... x
helidon => Roberta: You say must hear I first?
roberta => Helidon: You say hear it first?
helidon => Roberta: We were discussing you -- not me.
roberta => Helidon: We were discussing you -- -
helidon => Roberta: Oh, I ?
roberta => Helidon: Oh, and ?
helidon => Roberta: Please go on.
roberta => Helidon: Please move on.
helidon => Roberta: I'm all ears!
roberta => Helidon: I'm all set
helidon => Roberta: Can you elaborate on that?
roberta => Helidon: Can you on...
helidon => Roberta: What are your feelings now?
roberta => Helidon: What your options now?
helidon => Roberta: Why do you ask?
```

I'm not calling
these sensible
conversations.
LOL!



Yay! Puzzle time!

Puzzles: RoBERTa

- Did you notice anything unusual about the chatbots?
- What can we do to make it better?
- Can we use other models?
- Can we add more bots to the network?
- Can we make them make pass the Turing test?
- Is it a good idea?
- Any other ideas come to mind?



Demo: NLP Profiler

About the demo: NLP Profiler

- **Notebook & code on GitHub:**

https://github.com/neomatrix369/awesome-ai-ml-dl/blob/master/examples/better-nlp/notebooks/jupyter/nlp_profiler.ipynb
(https://github.com/neomatrix369/awesome-ai-ml-dl/blob/master/examples/better-nlp/library/org/neomatrix369/nlp_profiler.py)

- Simple python library-like module
- Liken to pandas-profiling but a lot simpler
- Tried to emulate Pandas' `describe()` function
- Can run on CLI (with changes) or `import` into a notebook
- Illustrates some use cases using a simple dataset
- More features can be built on top of the existing ones
- Can be integrated into the Better NLP library

nlp_profiler.ipynb
nlp_profiler-granular.ipynb
better_nlp_summarisers.ipynb
better_nlp_spacy_texacy_ex

Python 3

```
[5]: profiled_text_dataframe = apply_text_profiling(text_dataframe, 'text')
profiled_text_dataframe
```

	text	sentiment_polarity_score	sentiment_polarity	sentiment_subjectivity_score	sentiment_subjectivity	spellcheck_score
0	I love 🏀 very much 😊.	0.380000	Positive	0.43	Objective/subjective	1.000000
1	2833047 people live in this area. It is not a ...	-0.106818	Negative	0.55	Objective/subjective	0.968800
2	2833047 people live in this area.	0.136364	Positive	0.50	Objective/subjective	1.000000
3	the sentence doesn't seem to too many commas,...	0.375000	Positive	0.75	Pretty subjective	0.923880
4	Today's date is 04/28/2020 for format	0.000000	Neutral	0.00	Very objective	0.711510

We can make better use of these continuous values!

Fuzzy mapping of scores to human-readable terms

nlp_profiler.ipynb nlp_profiler-granular.ipynb better_nlp_summarisers.ipynr better_nlp_spacy_texacy_ex

Python 3

```
[5]: profiled_text_dataframe = apply_text_profiling(text_dataframe, 'text')
profiled_text_dataframe
```

	spellcheck_score	spelling_quality	sentences_count	characters_count	spaces_count	words_count	duplicates_count	chars_excl
	1.000000	Good	2	21	5	4	0	
	0.968802	Quite good	3	56	11	11	2	
	1.000000	Good						
	0.923887	Quite good						
	0.711513	Pretty good	2	64	8	9	0	

May not be very accurate (~70%)

See https://en.wikipedia.org/wiki/Words_of_estimative_probability - how we map the probability scores to English words

Puzzles: NLP Profiler

- What are the limitations of the NLP Profiler?
- What can we do to make it better?
- Can we make it more accurate?
- If we have scaling issues how do we tackle it?
- Any other ideas come to mind?

Yay! More
puzzles!



Throwing oneself at the deep-end

Build sample apps (similar to the demos), use simple ideas, improve, share, rinse and repeat!

How do I learn?

See previous talks &
Appendix section

Summary

Love this part. To
sum it all up!

- Timeline of my career progression
- Demos & projects
- Things I'm excited about in AI/ML/DL
- Components of AI/ML/DL
- Links to “Previous talks” and other “Learning Resources” to follow
- Masters & Grandmasters: thought exercises in the Appendix section

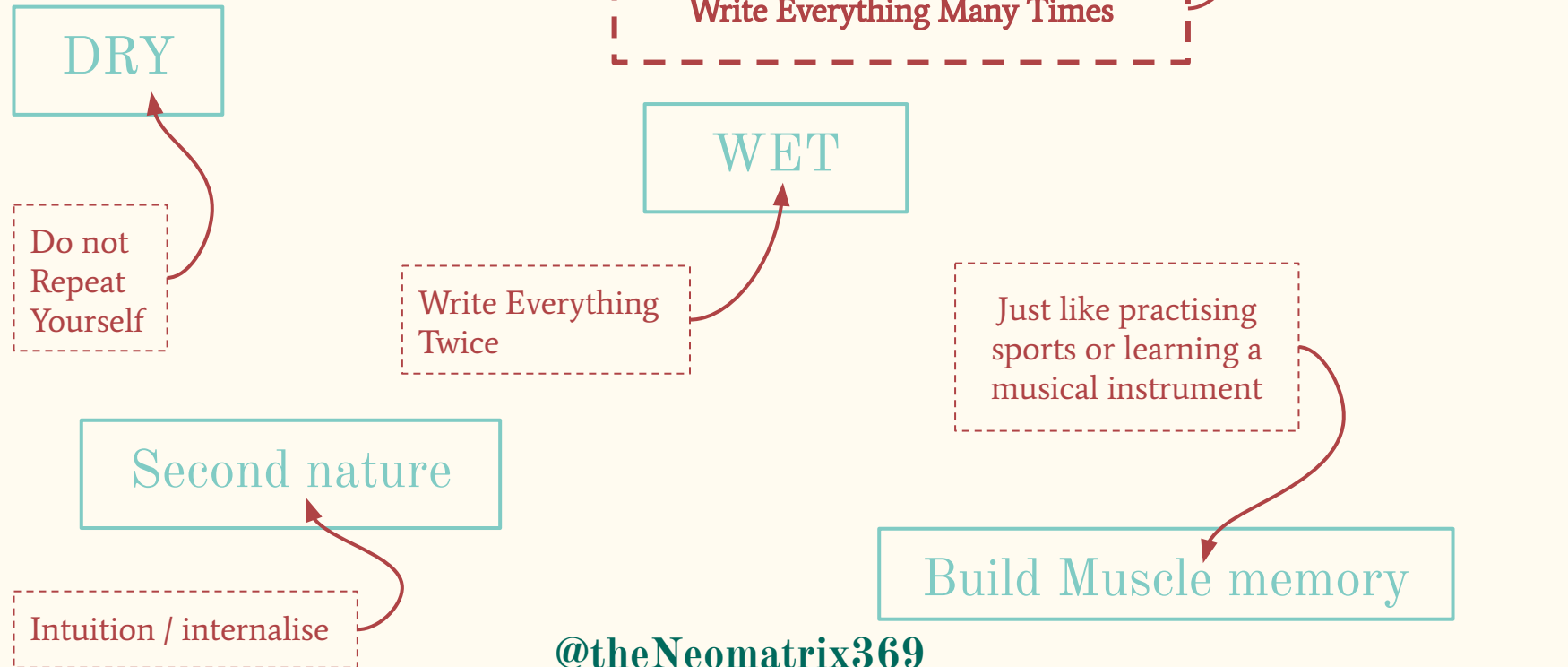


@theNeomatrix369

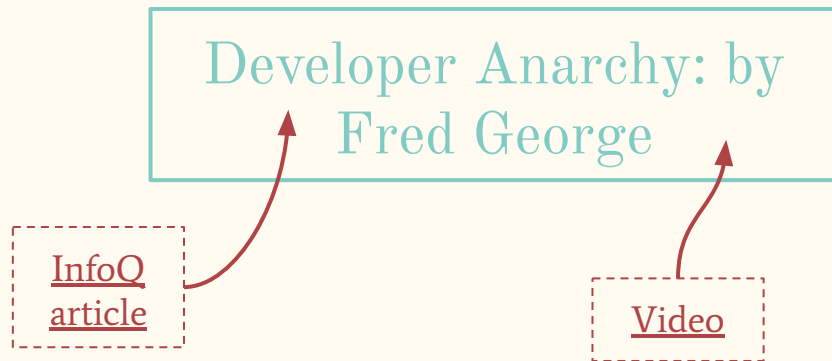
Previous talks

- I recently gave a talk on “nn” things every Java developer should know about AI/ML/DL
- Naturally, getting productive, my journey with Grakn and Graql
- Do we know our data as well as our tools?
- Java N.n: What to know? How to learn?
- Some of my other talks a can be found here and here (and others on Slideshare)


Repetition Theory



Repetition Theory: WEMT (Write Everything Many Times)



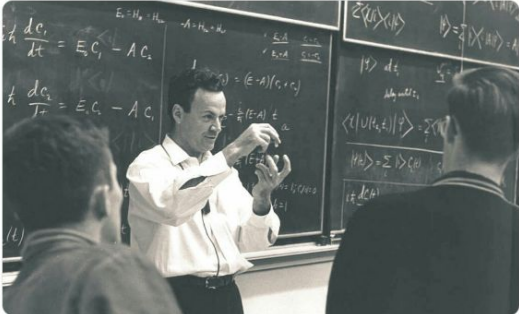
Repetition Theory: Richard Feynman quote

 **Richard Feynman**
@ProfFeynman [Follow](#)

The FEYNMAN method of learning:

- 1) Choose a concept and study it quickly
- 2) Teach it to someone, like a child, who is unfamiliar with the topic
- 3) Revise to fill in the gaps
- 4) Review and simplify

(repeat step 2 and 4 until you can explain the concept in your own words)



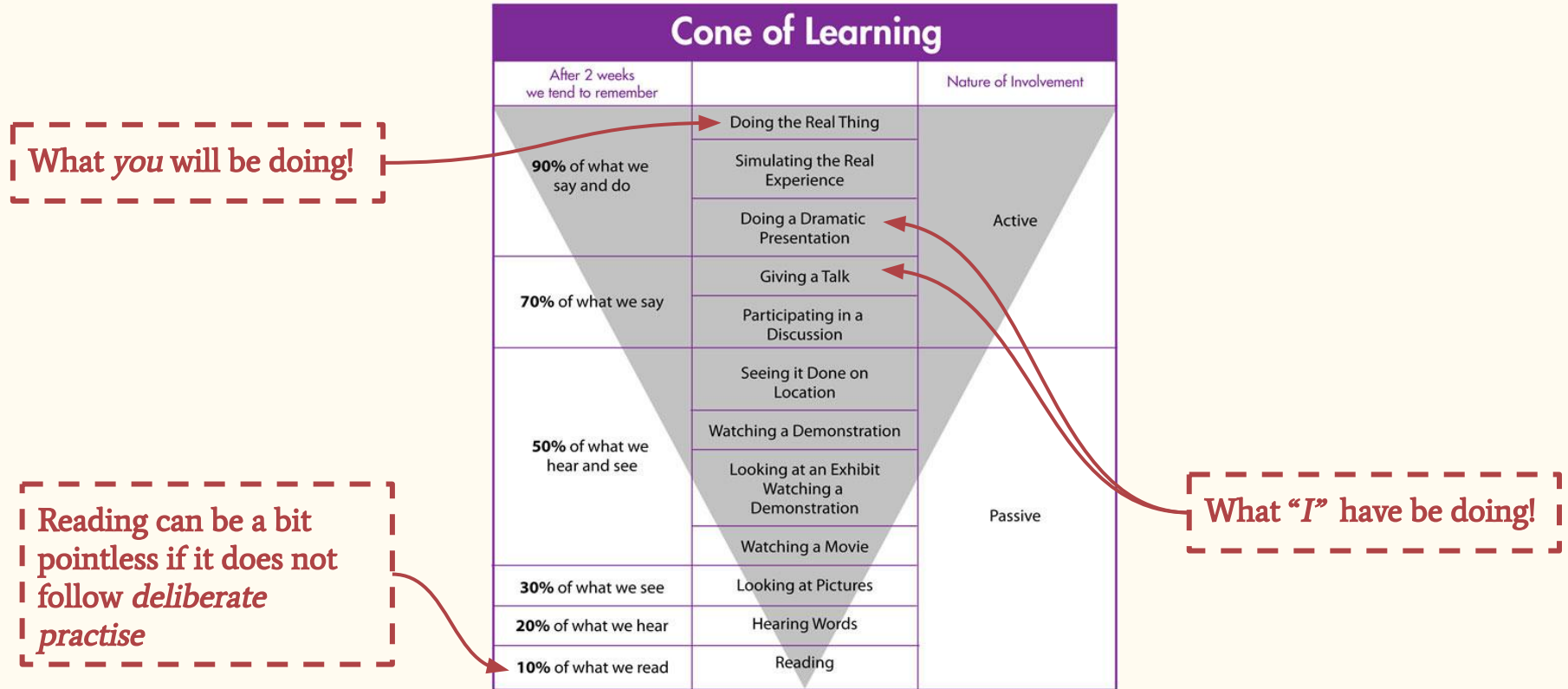
9:20 am · 25 Sep 2018

1,377 Retweets 3,369 Likes

20 1.4K 3.4K

theNeomatrix509

Cone of Learning



Insight

Writing code is Deep Work
Software Design is Deep Work

@theNeomatrix369

Learning websites

- [Awesome AI/ML/DL](#)
- [Better NLP](#)
- [Virgilio](#) | [GitHub](#)
- [MadeWithML](#)



All you need to know, pick and choose, make your own collections / checklists / playlists of resources to study and practice from.

Coherence with the brain

- Brain indexes knowledge
 - when you are resting
 - sleep and rest
 - let the brain or the mind index away
- Don't overload,
 - instead interleave
 - fill the gaps
 - learn incrementally & iteratively



Very important!!!

Book references

- Code Complete
- Clean Code
- Refactoring (Martin Fowler and Joshua Kerievsky)
- Four rules of simple design

Book references

- [Approaching \(Almost\) Any Machine Learning Problem](#)
- [How to think like a Scientist!](#)
- [Deep Work book](#)
- Getting Things Done: [Book](#) | [Free Resources](#)

Other Resources

So much to know and
learn, my, my!

- No need to feel overwhelmed
- Kagglers, read this for inspiration
- See Appendix section of the slide deck



Contact and keep in touch

Use Hashtag:
#abhishektalks

- twitter: **@theNeomatrix369**
- medium: **<https://medium.com/@neomatrix369>**
- github: **<https://github.com/neomatrix369/>**
- linkedin: **<https://www.linkedin.com/in/mani-sarkar/>**
- youtube: **channel | playlists**
- kaggle: **<https://kaggle.com/neomatrix369>**
- about me: **<https://neomatrix369.wordpress.com/about>**

Thank You!

Hey, tell us about
Abhishek's new book!

- **Abhishek Thakur**, for organising this session, and giving me a chance to present at this forum
- And to you, for sparing your valuable time and trusting in me



Closing

The pathway to mastery or championing a subject means we must be different and take different approaches, in addition to the ones that have already been taken!

Q & A

What is...

When is....
When to....

Why ...

Where can
I....

How to...

Can I ask the
“5 Whys”?

Who....

@theNeomatrix369

Appendix

Learning by example

@theNeomatrix369

DL4J example

- [Github](#)
- [Blog post](#)

NLP examples

- Example 1
 - Github
 - Blog post
- Example 2
 - Blog post
- Example 3
 - Blog post
- Better NLP

Jupyter Notebook example

- Example 1
 - Github
 - Blog: Exploring NLP concepts using Apache OpenNLP inside a Jupyter notebook
- Example 2
 - Blog post
- Example 3
 - Github
 - Blog post

Apache Zeppelin example

- Github
- Blog posts
 - Apache Zeppelin: stairway to notes* haven!
 - Running Apache Zeppelin on Oracle Cloud Infrastructure

graqL-to-english, english-to-graqL example

- Presentation
- Github

grCuda example

- Blog posts

- [grCUDA: A Polyglot Language Binding for CUDA in GraalVM. NVIDIA Developer Blog, November 2019.](#)
- [grCUDA: A Polyglot Language Binding. Presentation at Oracle CodeOne 2019, September 2019.](#)
- [Simplifying GPU Access. Presentation at NVIDIA GTC 2020, March 2020](#)
- [Optimizing Machine Learning Performance at Netsuite with GraalVM and NVIDIA GPUs](#)

- [Github](#)

graalPython examples

- Blog posts
 - [Introduction to the Python implementation for GraalVM](#)
 - [Moving from Jython to GraalVM](#)
 - [Running Python on GraalVM](#)
- [Github](#)

How do I learn?

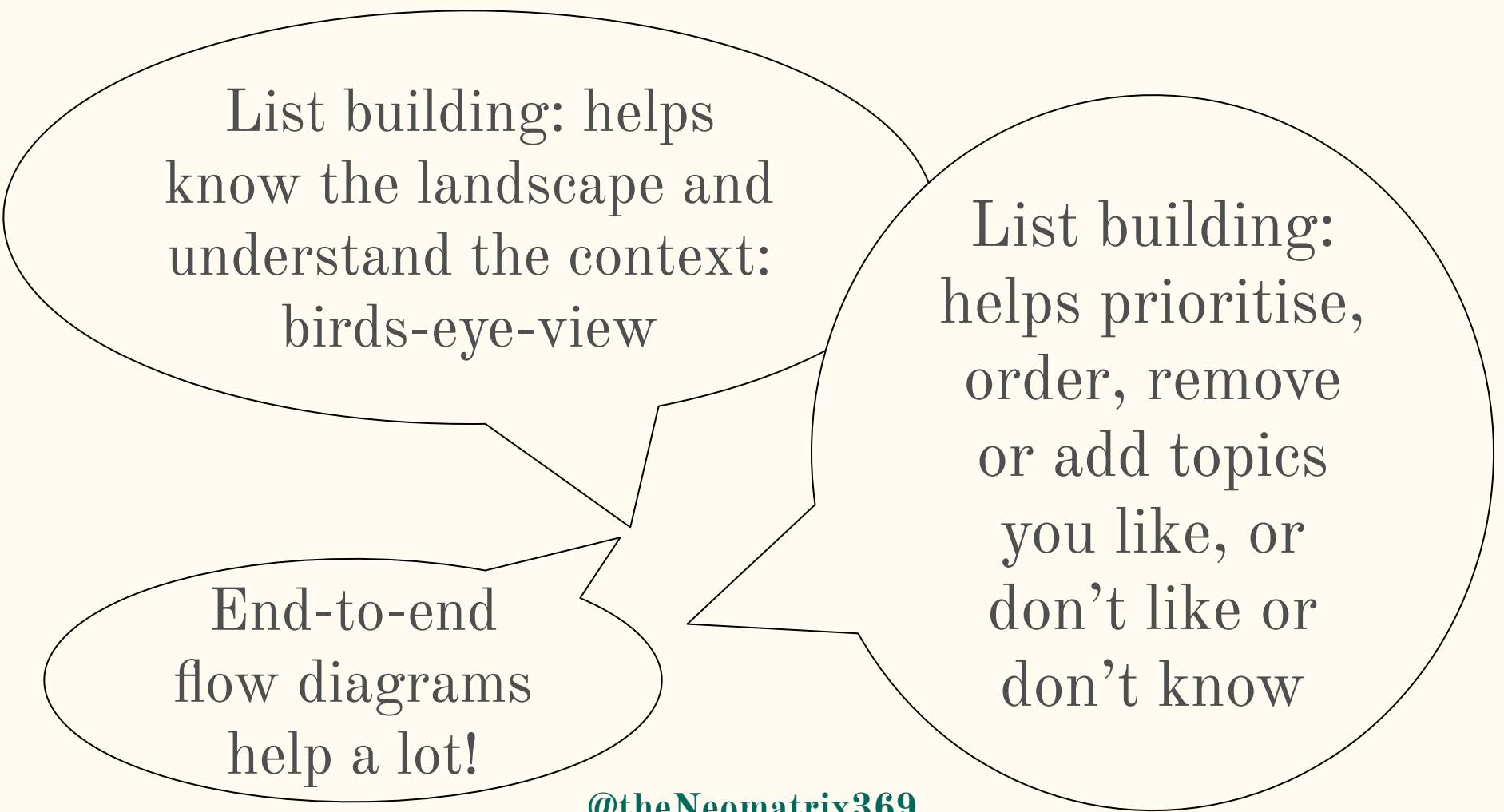
@theNeomatrix369

Top-down, outside in
learning, see Rachel
Thomas' talk
(fast.ai)

Mind-maps,
gathering reading
lists, organising
them, categorising
them

Nodes of Knowledge
(see Appendix,
previous talks)

Automatic
chunking, Learning
to Learn course by
Coursera



List building: helps
know the landscape and
understand the context:
birds-eye-view

End-to-end
flow diagrams
help a lot!

List building:
helps prioritise,
order, remove
or add topics
you like, or
don't like or
don't know

One may find these methods
unconventional or *non-mainstream*
but they do work and
give good results!

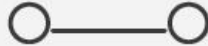
How to build and how
not to?

@theNeomatrix369

————— How **not to build** a minimum viable product —————



1



2



3



4

————— How **to build** a minimum viable product —————



1



2



3



4



5

Useful advice

@theNeomatrix369

See what Subhaditya says about intuition

- His talk on Abhishek's talk series: Talks # 2:
Subhaditya Mukherjee - Image restoration using
Deep Learning: Dehazing
- And this blog post on “imposter syndrome”

From KDM Delhi Kaggle Expert & Kaggle Master

- Sound advice from : Talks # 2: Subhadiya Mukhe1st Anniverary Meetup #3 || AMA with Mayank Kumar & Siddharth Yadavrjee

Masters and Grandmasters

@theNeomatrix369

Why do they inspire us?

- Start early: the amazing performance by little Souparnika Nair
- ...and keep going
- ...meet the masters (the droplets returns to the ocean)

*What's a common thread in all of them, in all of their performances?
(thought exercise)*

Master of grandmasters?

Britain's Got Talent 2020 Souparnika Nair: 10
Year Old Singer Full Audition S14E07

*What does it take to become such a talented person? Why?
(thought exercise)*

@theNeomatrix369

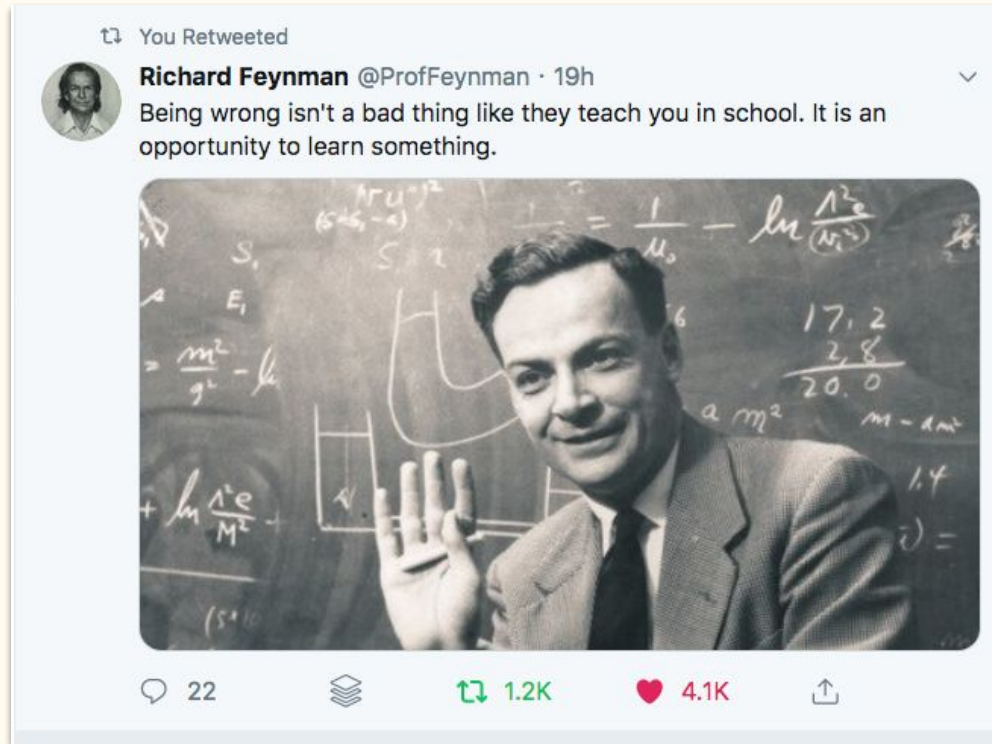
*The journey continues, the road maybe
long, but it's full of learnings that help
us grow - many a times at winding
points as the milestones keep changing.
We are living in such unprecedented
times!*

Feynman quotes


@theNeomatrix369

Being wrong isn't bad...

Me believe same!



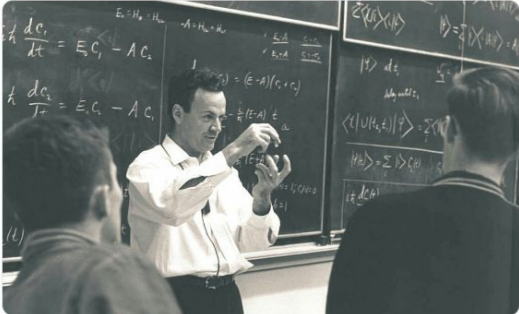
Repetition Theory: Richard Feynman quote

 **Richard Feynman**
@ProfFeynman [Follow](#)

The FEYNMAN method of learning:

- 1) Choose a concept and study it quickly
- 2) Teach it to someone, like a child, who is unfamiliar with the topic
- 3) Revise to fill in the gaps
- 4) Review and simplify

(repeat step 2 and 4 until you can explain the concept in your own words)



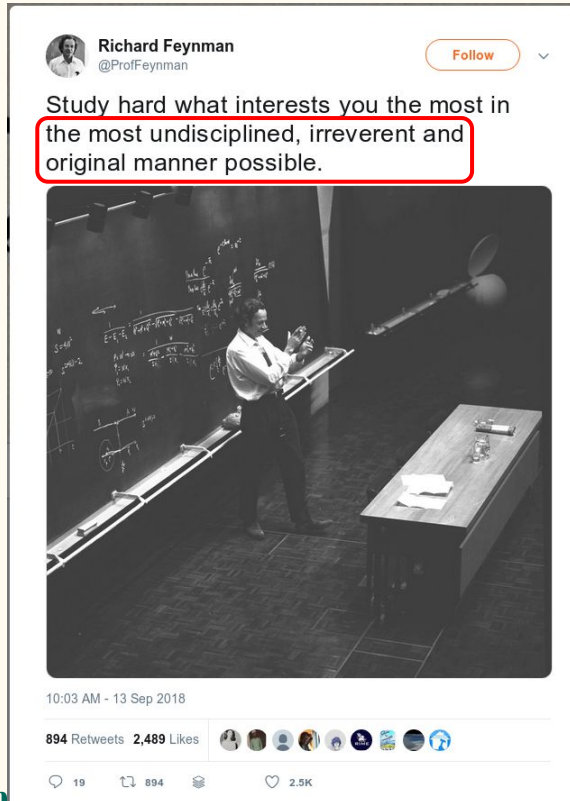
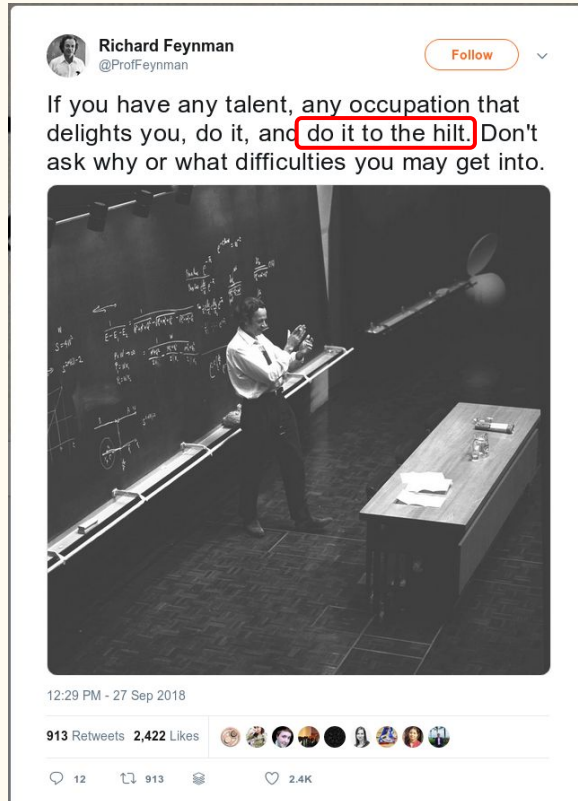
9:20 am · 25 Sep 2018

1,377 Retweets 3,369 Likes

20 1.4K 3.4K

theNeomatrix509

More Richard Feynman quotes...



Other fun stuff

@theNeomatrix369

Comedians really know their data!



55
seconds
of fun

[Link to the video](#)

@theNeomatrix369

Comedians really know their data!

Ellen D.
knew her
time-series

She made us
laugh, so she
must know
something...

Ellen D.
knew her
graphs



We were the
mascots!

Call us
sidekicks, if you
will!

It has been an
honour and a
pleasure to be
here!

