

Business Analyst?

Data Analyst?

Data Scientist?

Data Engineer?

ML Engineer?

Business Analytics?

Business Intelligence?

Data Science?

Artificial Intelligence?

Machine Learning?

Your Introduction to Data Science and AI



JCOp Untuk Indonesia

Part 1/3

Data Science

Data Science (Dhar, 2013)

- Interdisciplinary field
- Extract insight from data



How to leverage your data

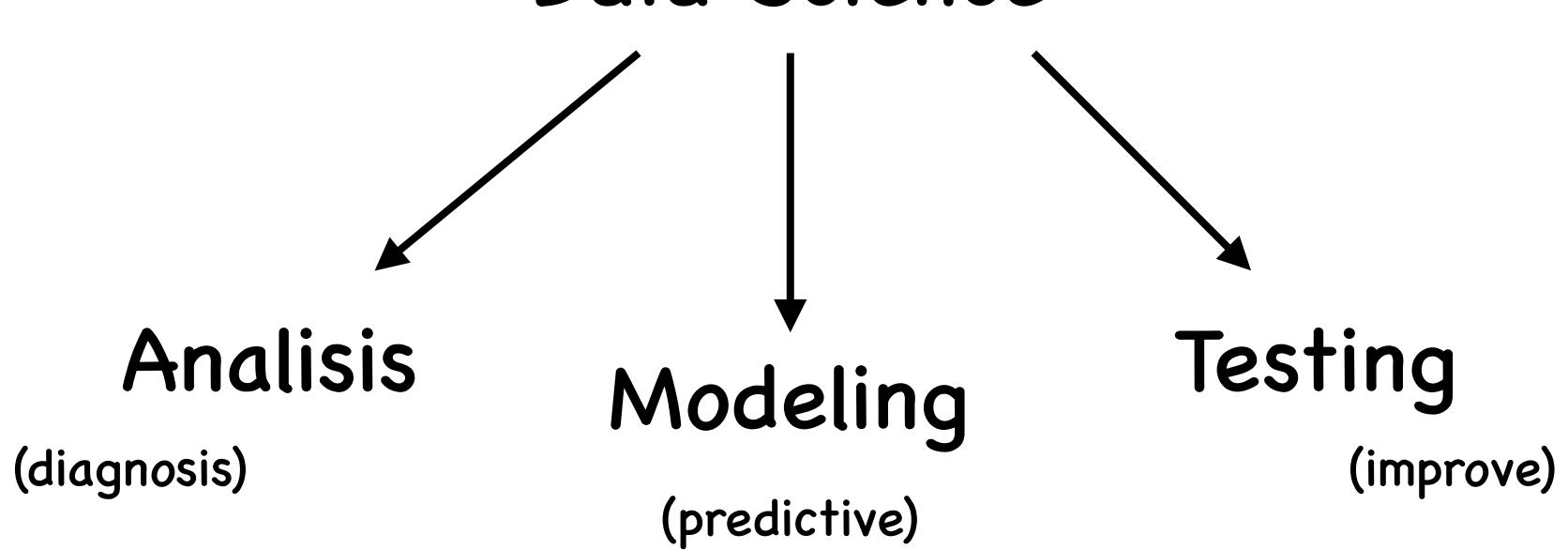
Descriptive Analytics
(Rangkuman / Laporan)

Diagnostic Analytics
(Analisa / Penyebab)

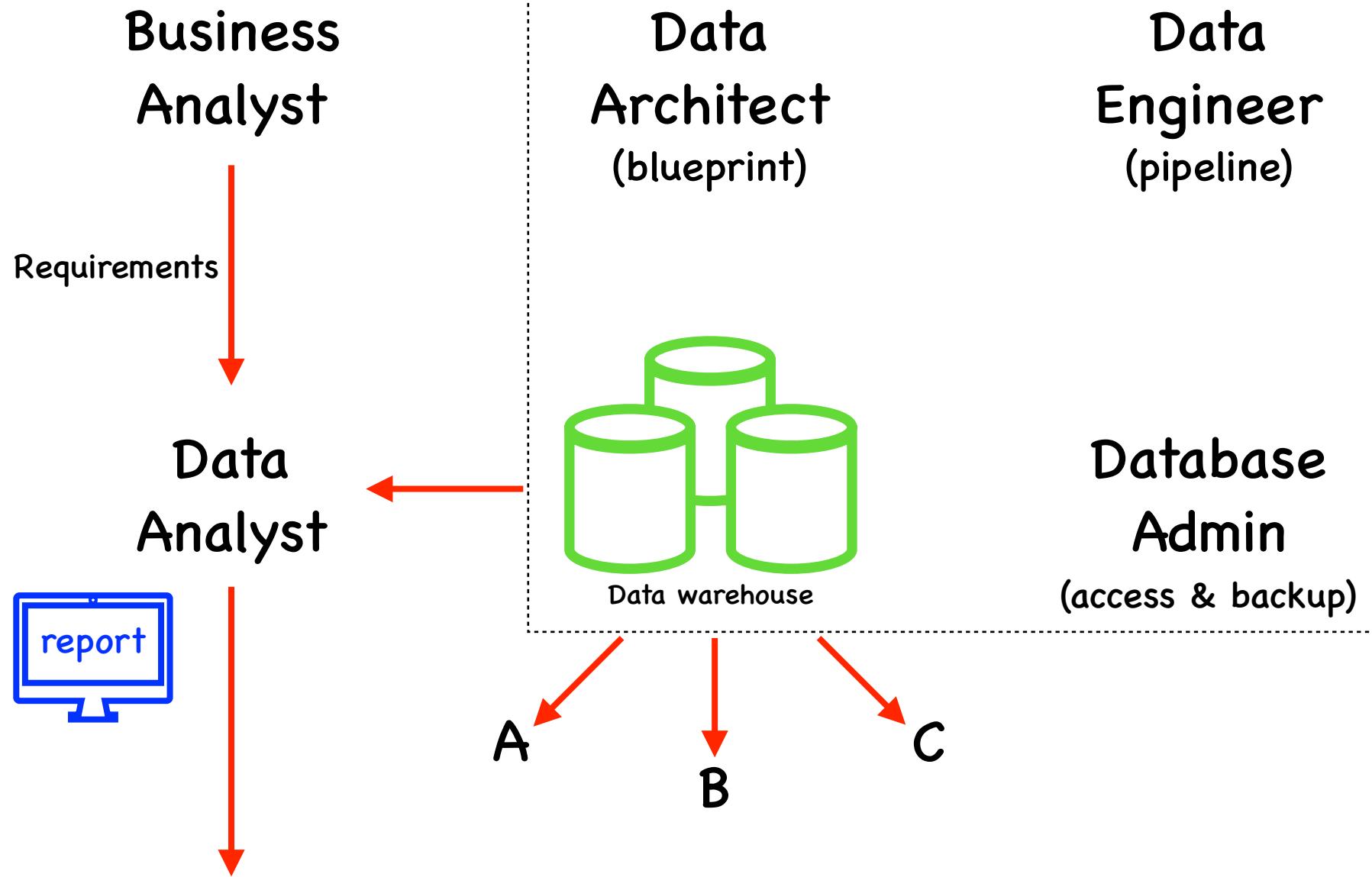
Predictive Analytics
(Prediksi)

Prescriptive Analytics
(Rekomendasi / Aksi)

Data Science

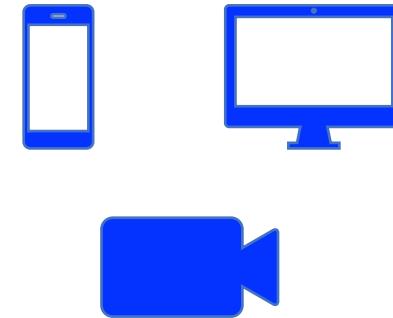


Analisis



Modeling

Data Scientist → ML Engineer
AI Engineer



Back End Developer

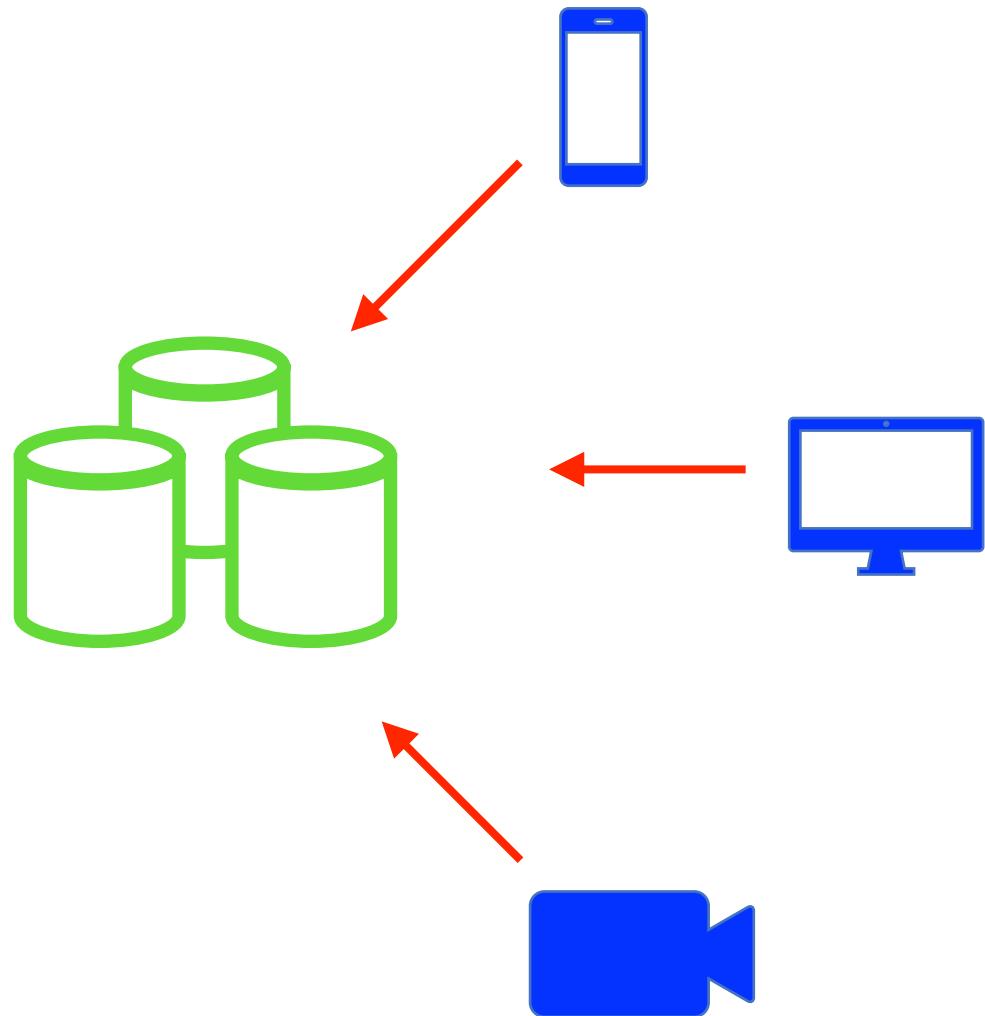
Front End Developer

DevOps Engineer

(Mobile Developer)
(Web Developer)
(Cloud Engineer)

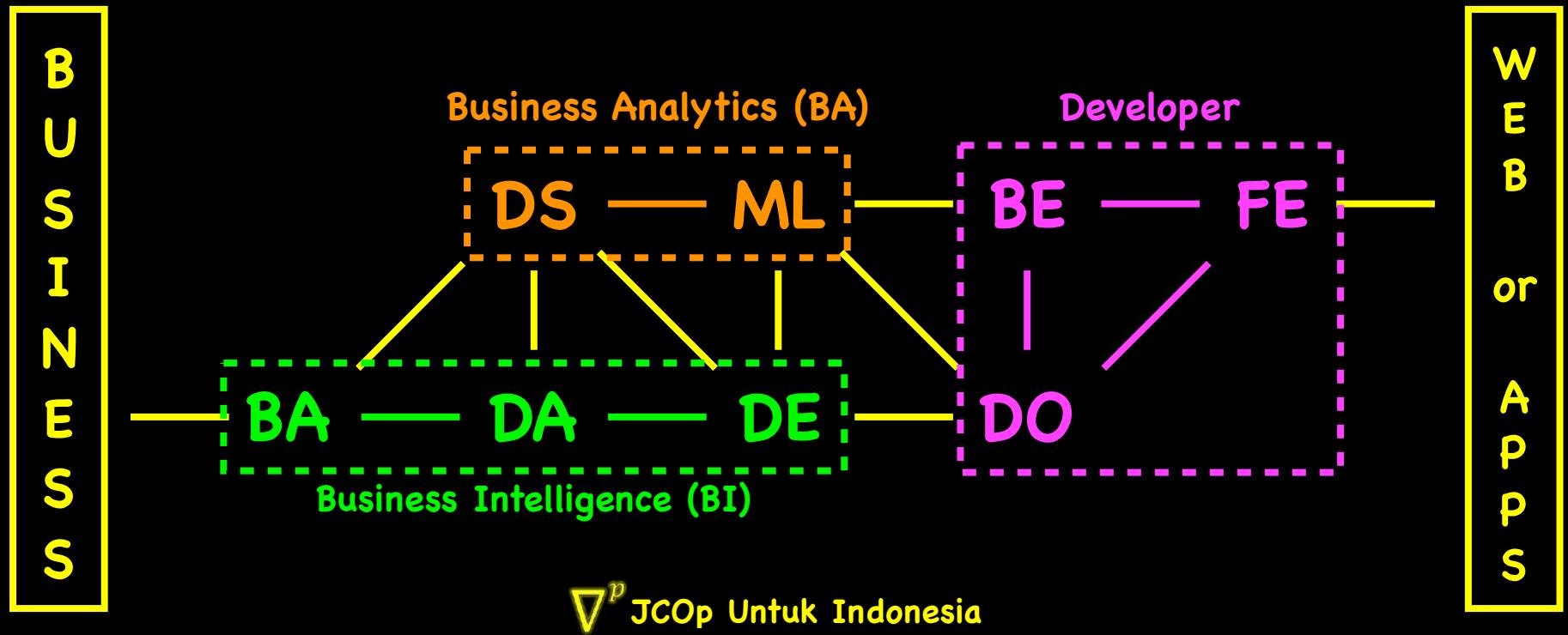
Testing

Data
Scientist





∇^p JCOp Untuk Indonesia



Using AI for Data Science problem?
Here comes Machine Learning

Note: It doesn't have to use AI though

Part 2/3

Artificial Intelligence
dan
Machine Learning

Artificial Intelligence

- **Simulation** of Natural Intelligence in machines
- Rationalizes input
- Take action

Machine Learning (Arthur, 1959)

- A field of study
- Gives machine the **ability to learn**
- **without explicitly** programmed

Artificial Intelligence

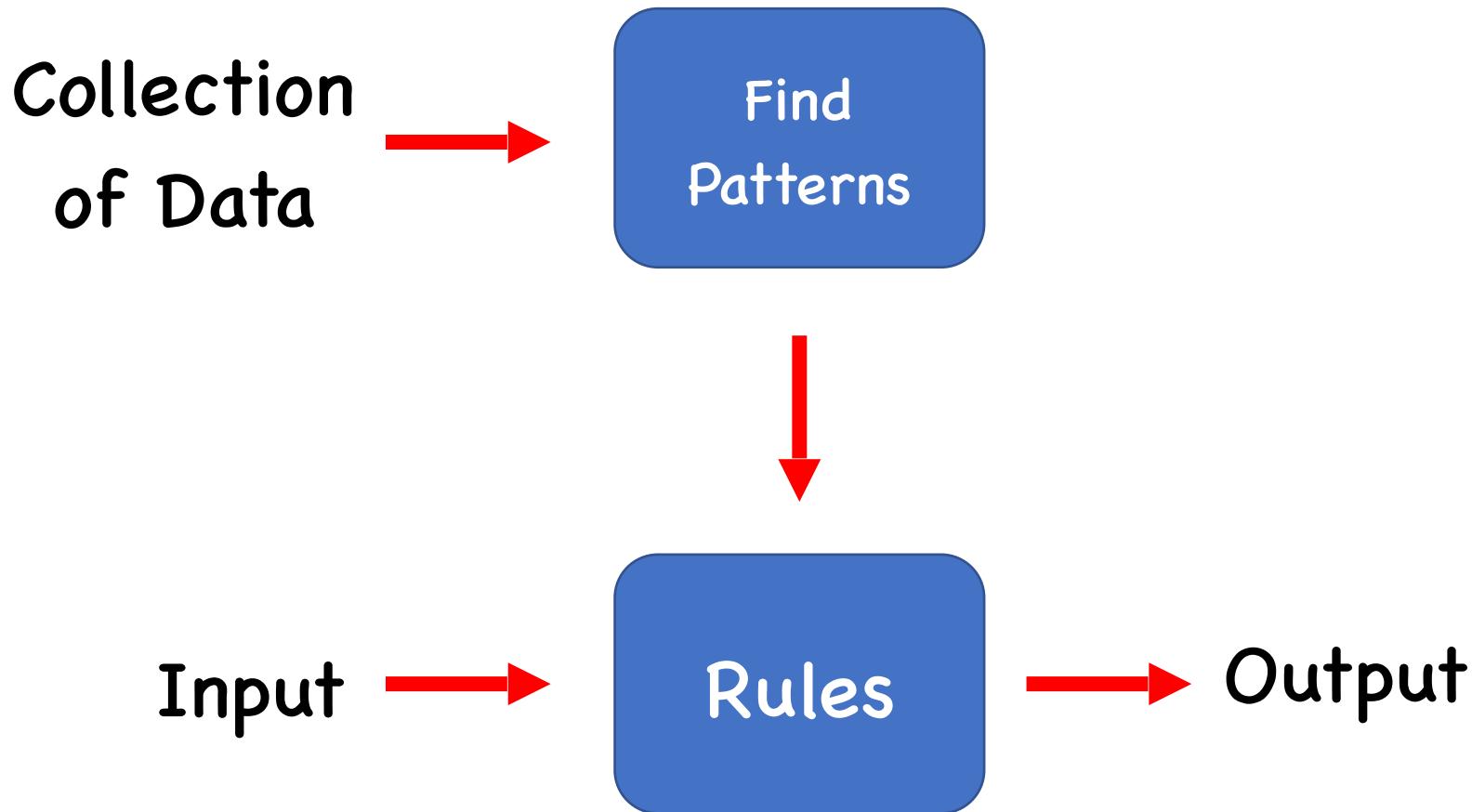
Expert

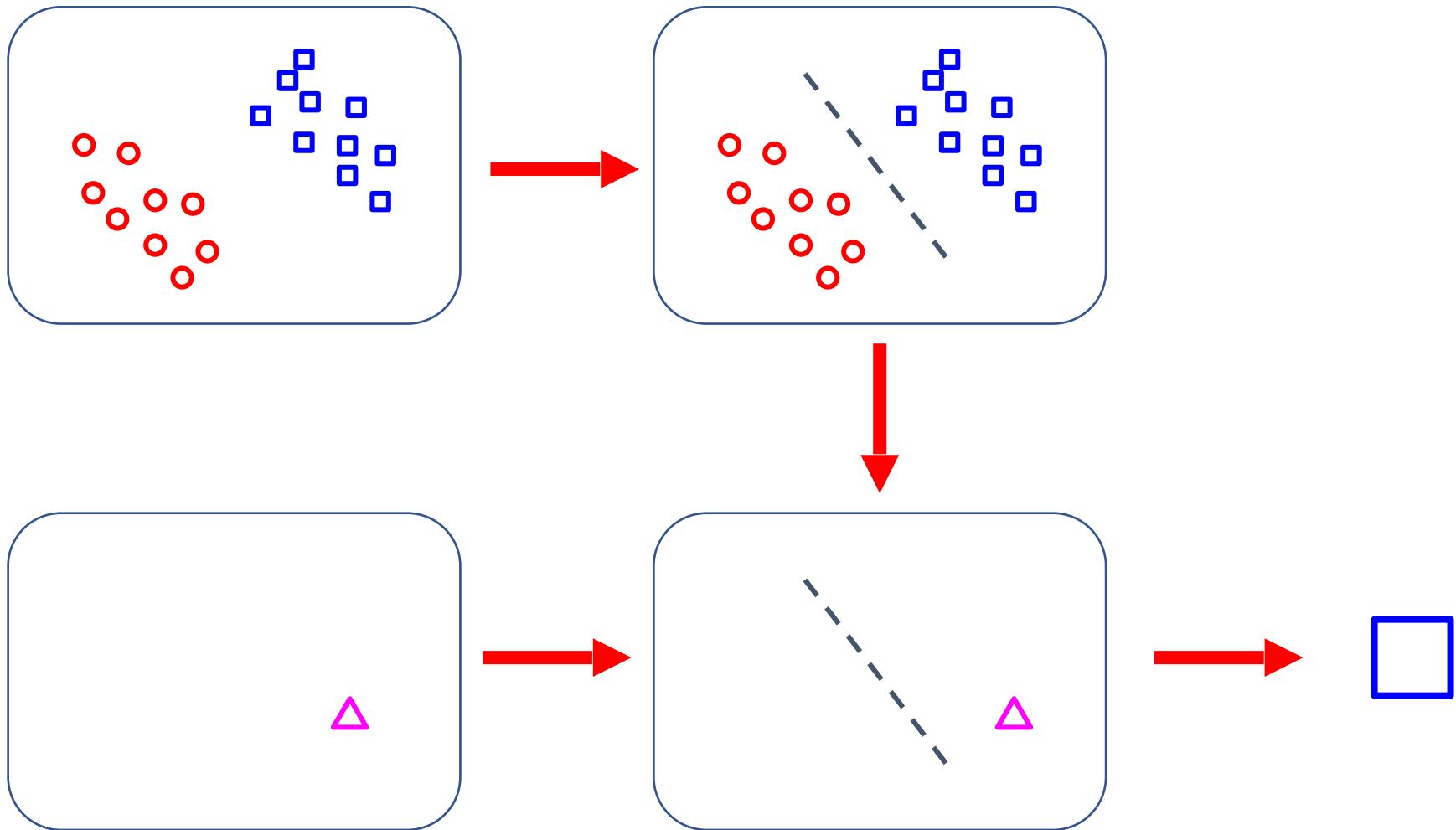


programmed
explicitly

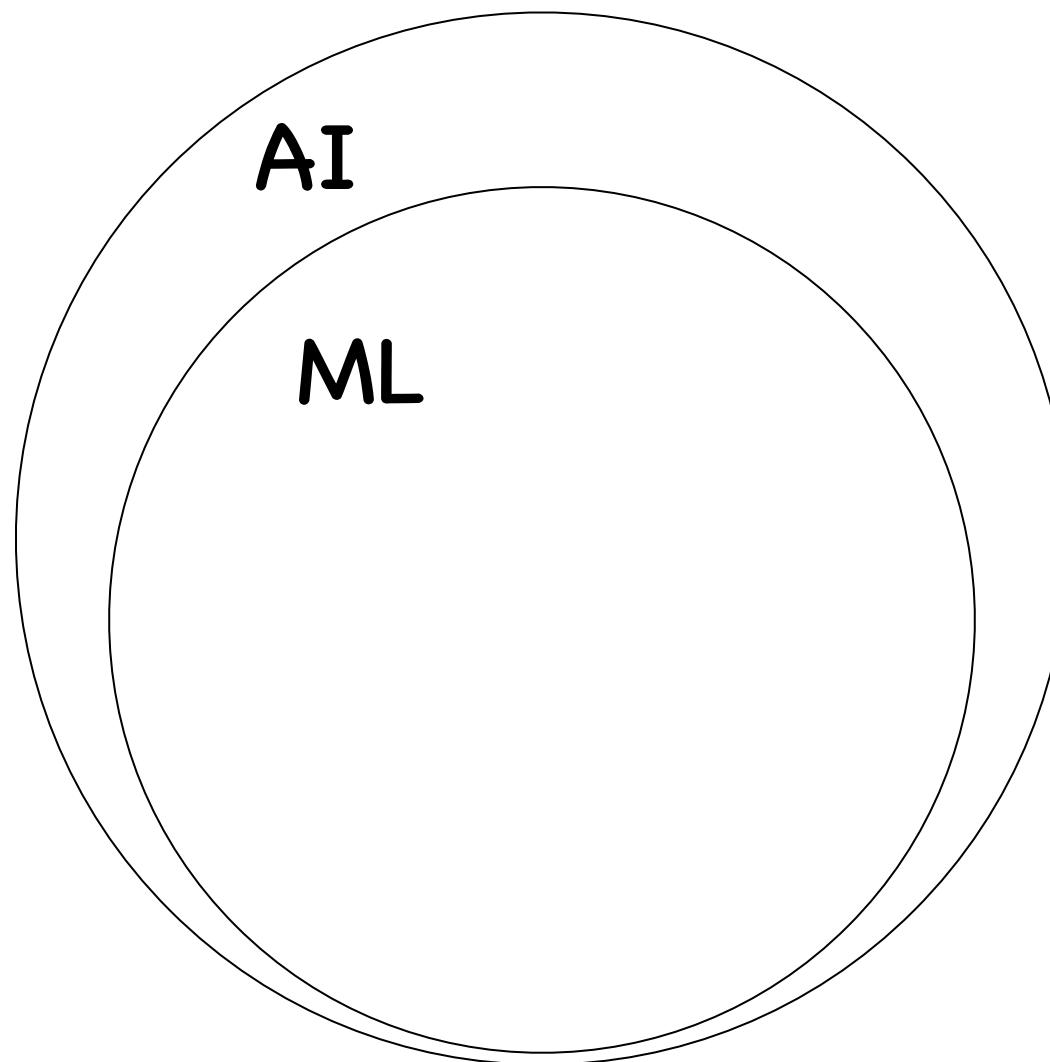


Machine Learning

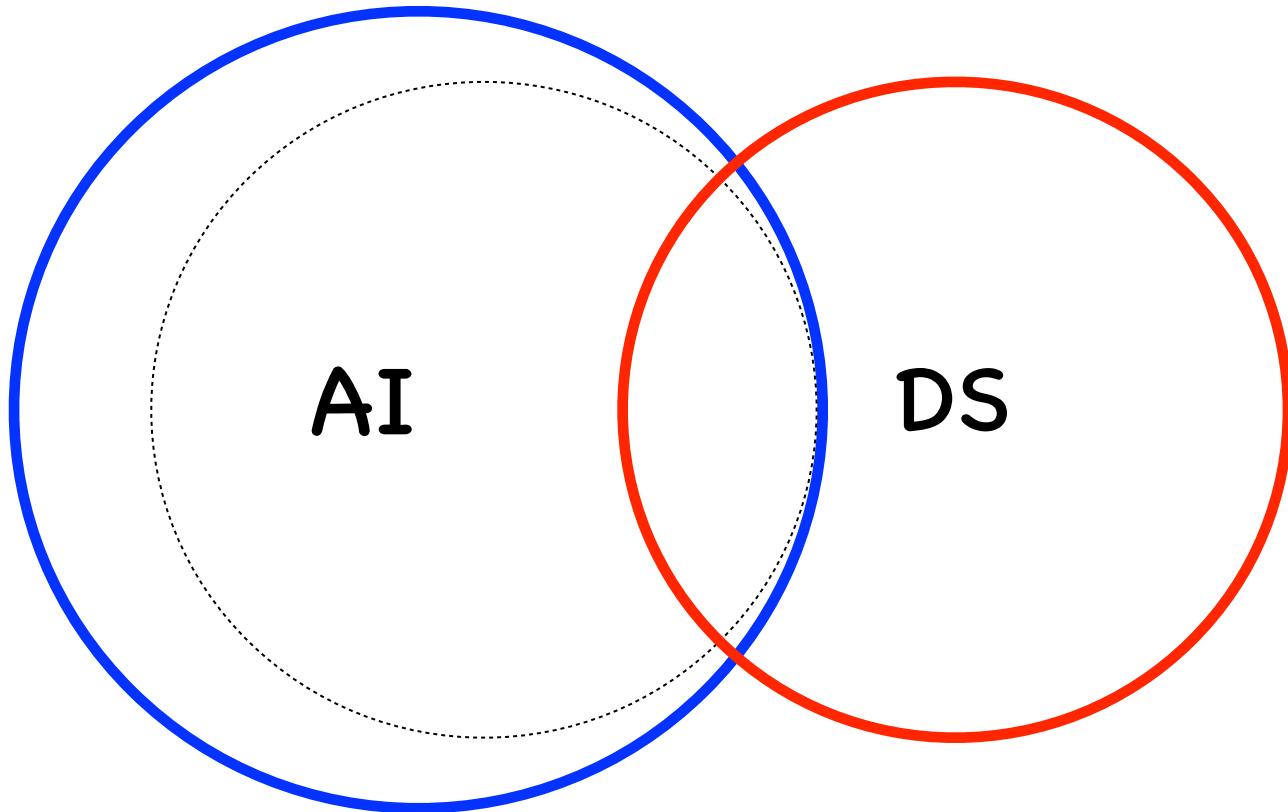




AI vs ML

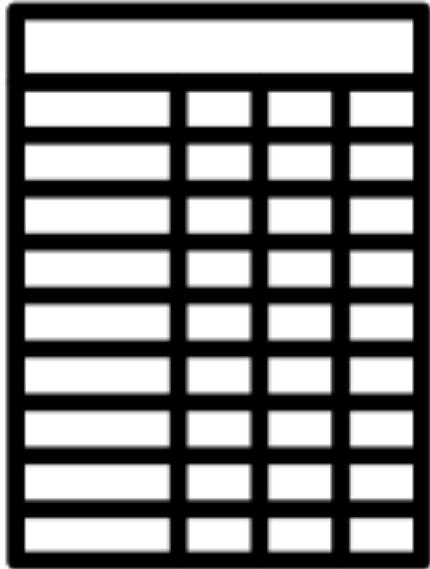


AI vs ML vs DS



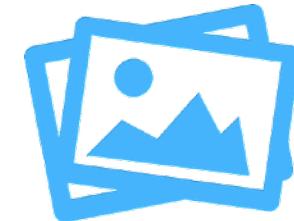
Evolution of Data

Previously



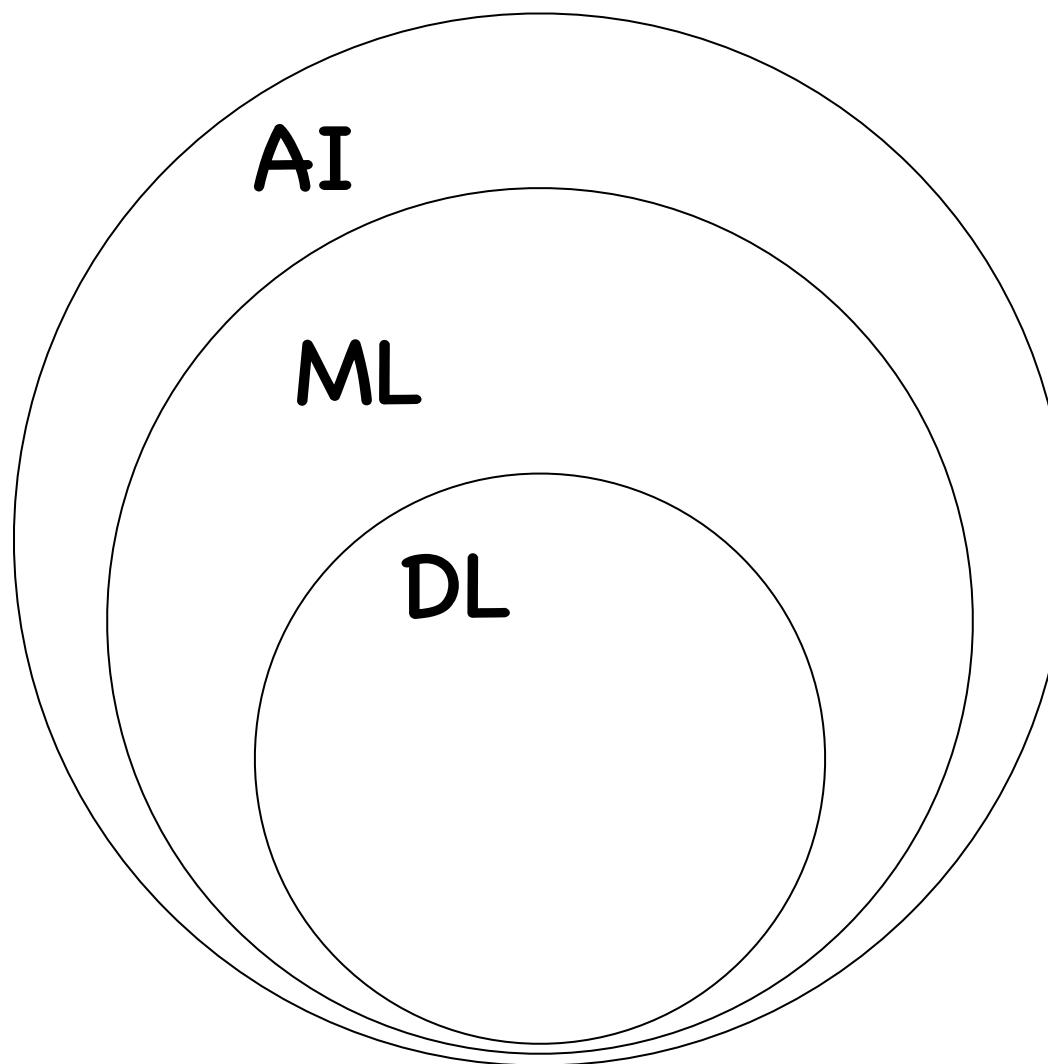
Structured Data
(ML is good at this)

Recently



Unstructured Data
(DL is good at this)

AI vs ML vs DL



Deep Learning pushes the limit

6 Human Mental Task

Reasoning
Knowledge Representation
Planning and Navigation
Natural Language Processing
Perception
Generalized Intelligence

Better Reasoning

Chess in the AI era (1985)

Kasparov vs 32 best chess machine at the same time



Kasparov won all the games

The Brain's Last Stand (1997)

Kasparov vs Deep Blue (IBM)



$2\frac{1}{2} - 3\frac{1}{2}$... Deep Blue won the game



Stockfish



Komodo



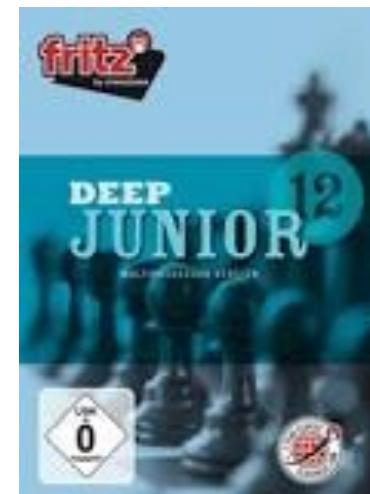
Shredder



Rybka



HIARCS



Junior

Chess in the era of Machine Learning



100 games: 28 win, 72 draw, 0 loss
4 hours training from scratch

Go



10^{172} possible position
There are 10^{80} atoms in the universe

AlphaGo vs Lee Sedol (2016)



4 - 1

Lee Sedol comments

- Game 3:

"I misjudged the capabilities of AlphaGo and felt powerless."

- Game 4 victory

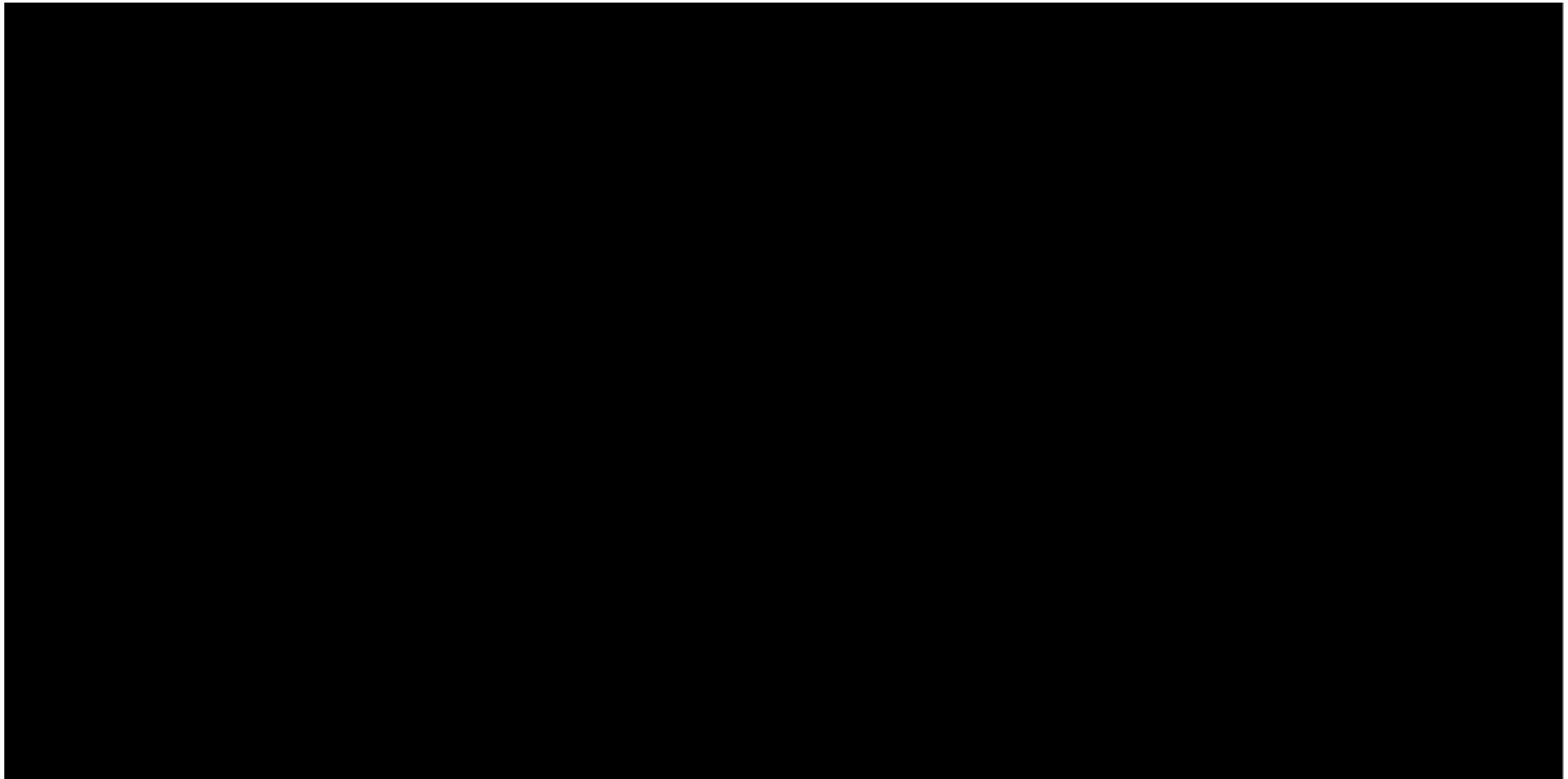
"priceless win that I (would) not exchange for anything."

- After the game:

"I've improved already. It has given me new ideas"

update: It is sad to hear that Lee Sedol has retired from the game of Go

Simple Snake Game



source: [Snake AI Fight](#)

Better Knowledge Representation

Deep Learning in Radiology

NORMAL



PNEUMONIA



Test accuracy: 0.9326923076923077

And more



NEURAFARM

Better at Planning and Navigating

DARPA Driverless Car Challenge

2004

- 240 km route in Mojave Desert, US.
- No winner.
- CMU went farthest: 12 km

2005

- 212 km route
- 1st : Stanford (6 hr 54 min)
- 2nd : CMU (7 hr 5 min)



DARPA Driverless Car Challenge

2007

- Urban challenge. 96 km in less than 6 hr.
- Must obey all **traffic regulations**
- 1st : CMU (4 hr 10 min) with 22.5 km/h in average



Autonomous Driving Programme

- Thousands of engineers
- Forbes, Tesla, Google, Toyota, Uber, BMW, etc.

Then a challengers come

George Francis Hotz



First hacker to Jailbreak iPhone at 17 y.o.
Jailbreak PS3

- Invited by friend to build technology for Tesla
- Geohot see what Tesla did and he is confident he can do better



Meet the 26-Year-Old Hacker Who Built a
Self-Driving Car... in His Garage... in a month

Note: This may be old news, but it shows the hype of deep learning at the time

Better NLU and NLP

(Natural Language Understanding and Processing)

Named Entity Recognition

Paris Hilton was in Paris at the Paris Hilton

PERSON GPE ORG

Named Entity Recognition

Halo, nama saya **Wira**, dan saya sedang mencari **laptop macbook pro**. Saya sedang bekerja di **jakpus**.

```
[('Wira', 'NAMA'),  
 ('laptop', 'KATEGORI'),  
 ('macbook pro', 'PRODUK'),  
 ('jakpus', 'GEOPOLITIK')]
```

Autocomplete

Indonesia is home to the largest concentration of sea turtles in the world. The vast majority of these animals are protected by law and are protected by the Indonesian government . They are also known as the "Sea Turtle Defenders ", because it is their role to save the sea turtles that are being used for human consumption. The Sea Turtle Defenders , through their educational activities and their work with local communities , have saved the lives of thousands of sea turtles|

Better Perception

2D images to 3D models



source: [Hierarchical Surface Prediction for 3D Object Reconstruction](#)

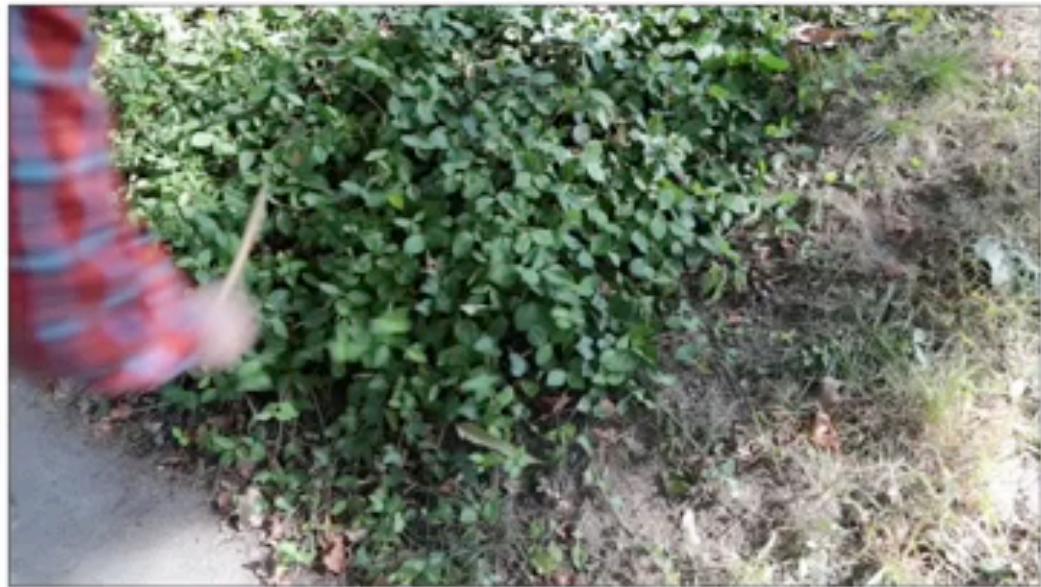
2D images to 3D models



Could be used in movies or games



source: [Visually-indicated sounds](#)



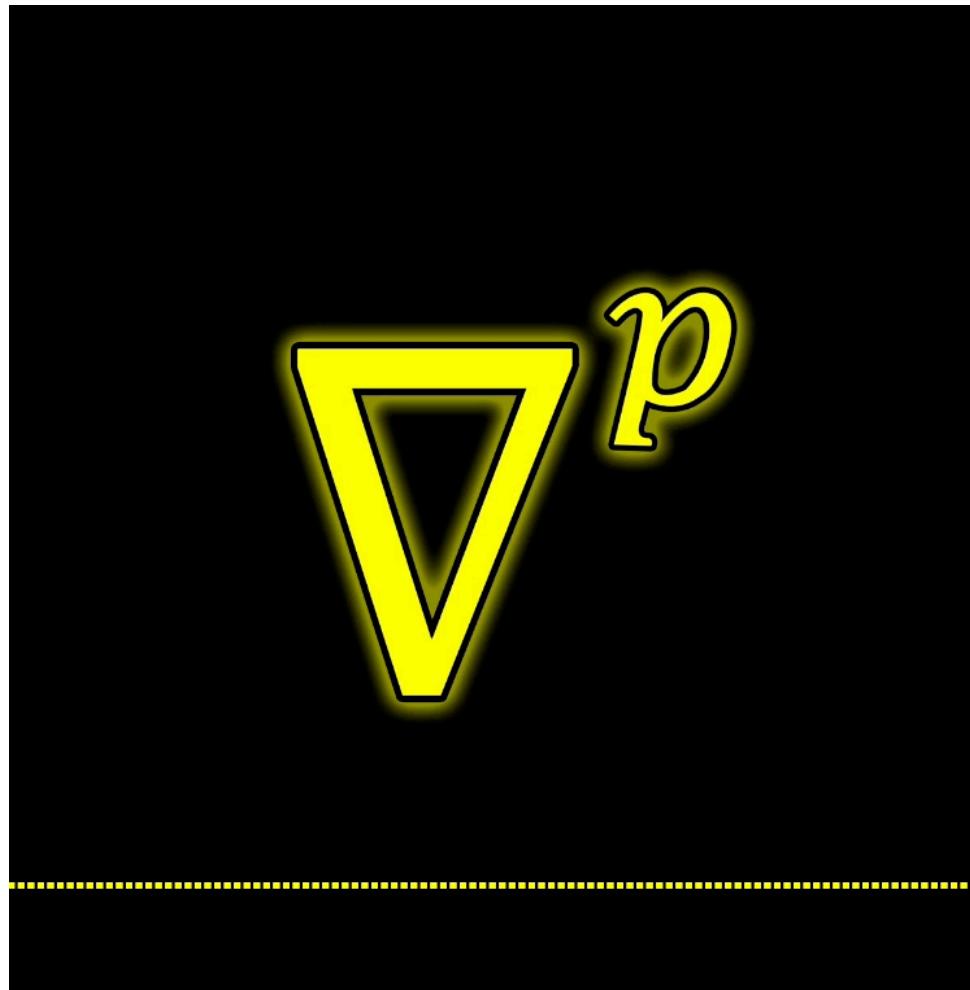
Input video



Transferred audio clips

source: [Visually-indicated sounds](#)

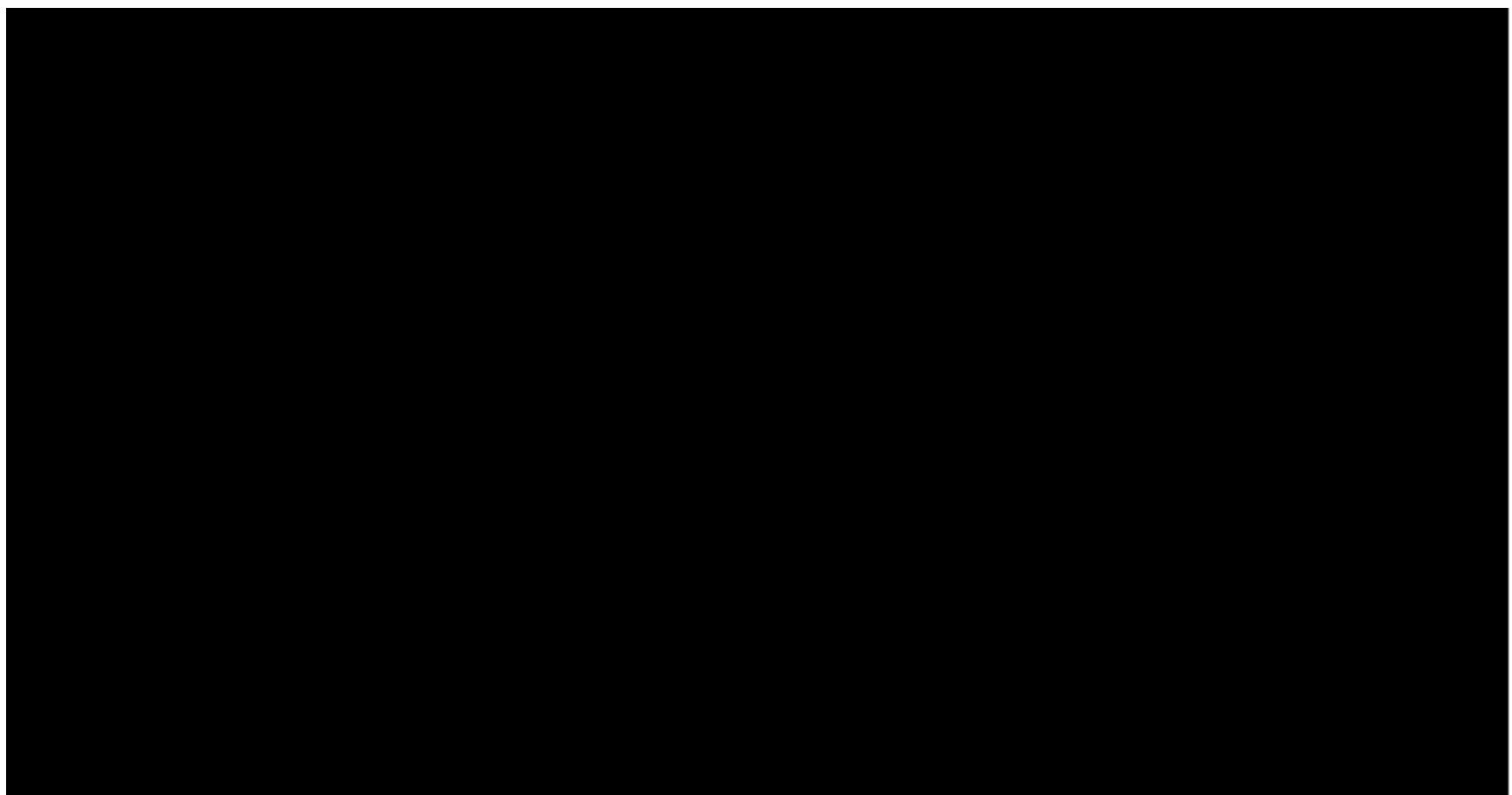
Text to Speech



source: [Text to Speech Bahasa Indonesia](#)

More Generalized Intelligence
(emotional intelligence, creativity, etc.)

MIDI Music Composition



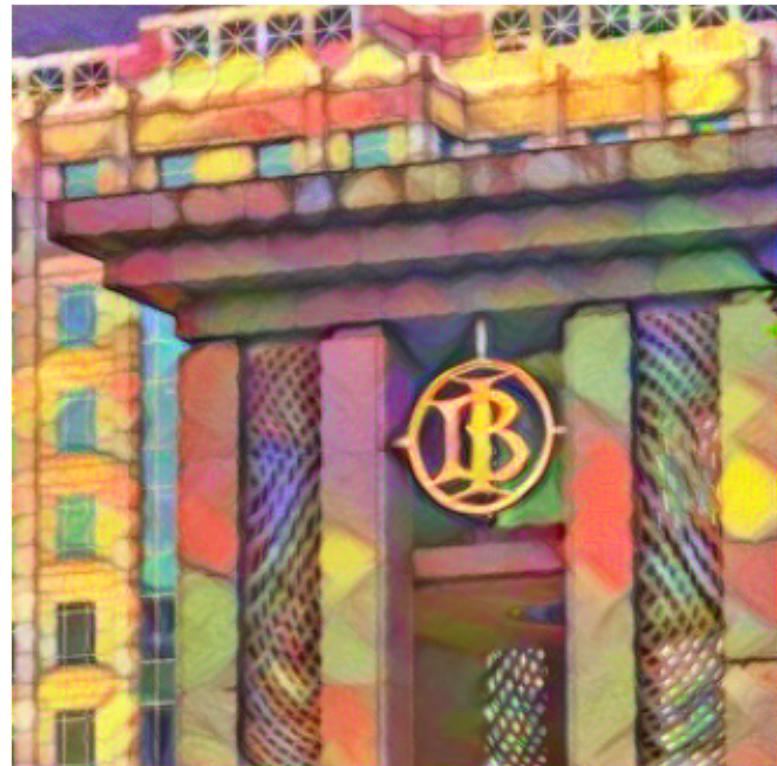
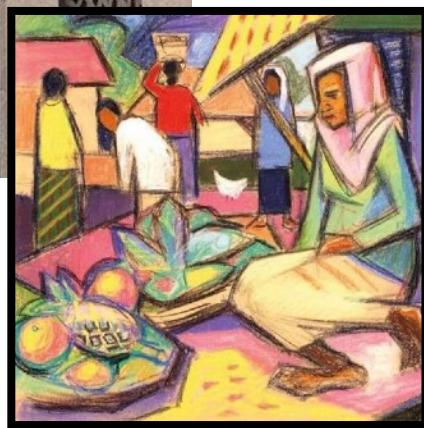
source: [MIDI Music Composition with Deep Learning – Code in Python](#)

Trained using 47 Indonesian Folk Songs

Neural Style Transfer



Try make your own [Neural Style Transfer](#)



Fruit Seller

Oleh
Barli Sasmitawinata

Deep Learning in daily life

Face detection and recognition



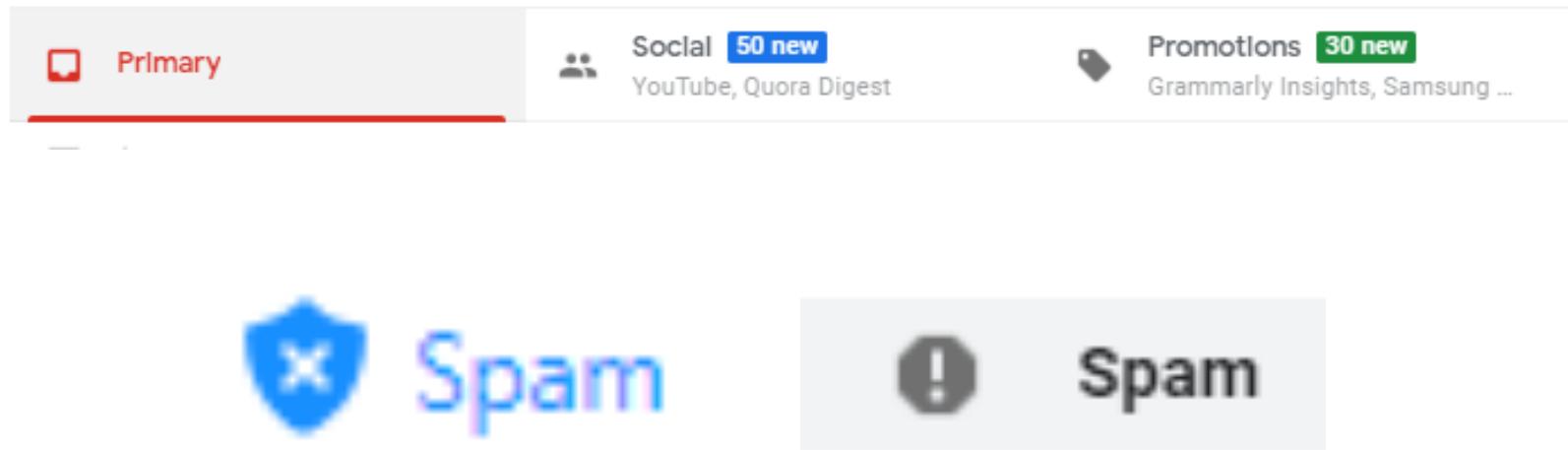
source: [Hendrik Tio Face Recognition](#)

Recommendation System

The screenshot shows the Spotify Premium desktop application. The main area displays the "Discover Weekly" playlist, which is described as a weekly mixtape of fresh music. It was created by Spotify and contains 30 songs, totaling 1 hour and 59 minutes. The playlist is currently set to "FOLLOWING". Below the playlist, there is a table listing the top songs from the Discover Weekly playlist. The table includes columns for Song, Artist, and Last Played (2 days ago). The songs listed are: Stardust - Extended Mix by Jay Hardway, Me, Myself & I - Viceroy Remix by G-Eazy, Bebe Rexha, ... Show Me Love 2K16 by Robin S., Chris Mont... Every Time (feat. Grant Genske) by Francis Davila, Alan ... Feel My Heart (Beating) - Radio Edit by David Pop, and Copacabana - Radio Edit by Nari & Milani, Hiisak. At the bottom of the screen, the playback controls show that the song "This Girl" by Harris, The Cooking Girls is currently playing, with a progress bar at 0:02 of 3:15.

SONG	ARTIST	LAST PLAYED
Stardust - Extended Mix	Jay Hardway	2 days ago
Me, Myself & I - Viceroy Remix	G-Eazy, Bebe Rexha, ...	2 days ago
Show Me Love 2K16	Robin S., Chris Mont...	2 days ago
Every Time (feat. Grant Genske)	Francis Davila, Alan ...	2 days ago
Feel My Heart (Beating) - Radio Edit	David Pop	2 days ago
Copacabana - Radio Edit	Nari & Milani, Hiisak	2 days ago

Email Classification



Face translation app



source: [Faceswap with DFL Styled Autoencoder](#)

Face translation app



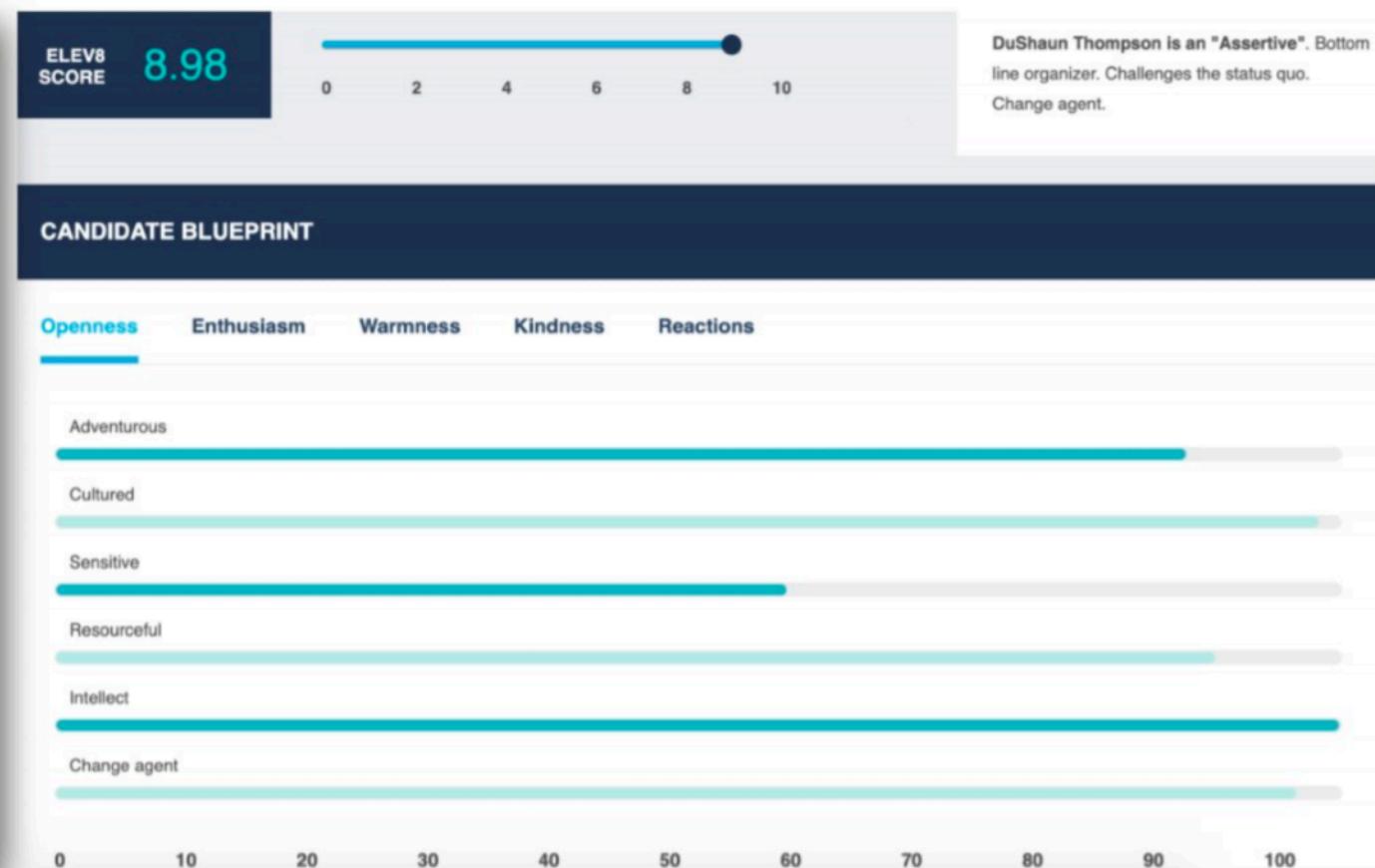
source: [CycleGAN](#)

Walk-out shopping (Amazon Go)



Recruitment Screening

Assess job suitability from 30s video

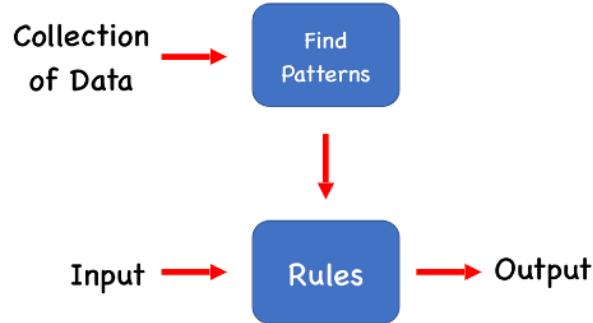


Too good to be true?

Part 3/3

Machine Learning Limitation

Machine Learning



Machine learning is a powerful pattern recognition tool, so it relies on

- pattern exists and not bias
- pattern can be extrapolated beyond data

Extrapolation doesn't always work



Check your spam folder

Comments (7)

findjob in North Carolina
74 months ago

Earlier I happened to check my email in my spam folder. I was surprised to find a rejection email for a job I applied for in August. I was wondering how many other rejection emails I may have "missed" as a result of the email going to my spam folder.

So just a friendly reminder to check your spam emails. Many jobs we apply for and don't hear anything back...the company may have very well sent an email and it just happened to go in your spam folder.

[Reply](#)

 Michael Margolis
@yipe



Alexa: remind me to feed the baby



287K 4:02 AM - Jun 10, 2018



73.4K people are talking about this





Parametric synthesis





Deirdre Bosa

@dee_bosa



I think I just shoplifted?? [#AmazonGo](#) didn't charge me for my Siggi's yogurt 😬 [#nolinesnocheckout](#) [#freestuff](#)



Amazon Go Vacuum Insulated
Bottle, 17 oz

\$14.99



Essential Baking Company
Oatmeal Raisin Cookie, 80g

\$2.29
\$2.29

Subtotal \$28.55

Estimated Tax \$2.11

Total \$30.66

♡ 275 10:17 PM - Jan 22, 2018



126 people are talking about this



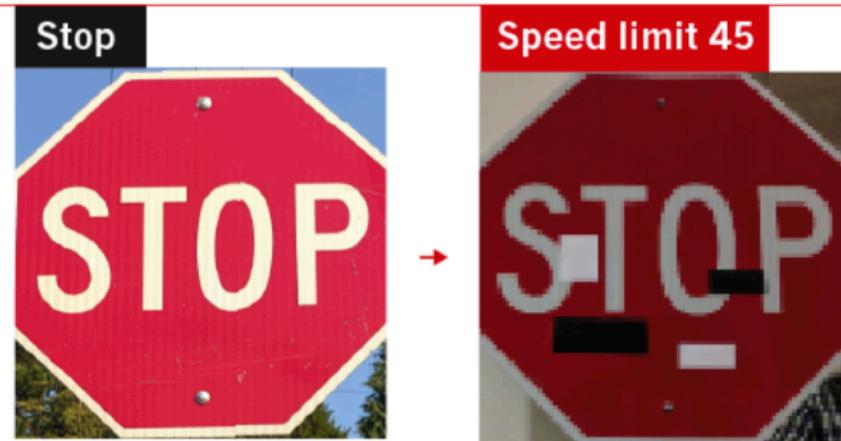
Note: It costs on avg \$1M/store, BEP ~2 years



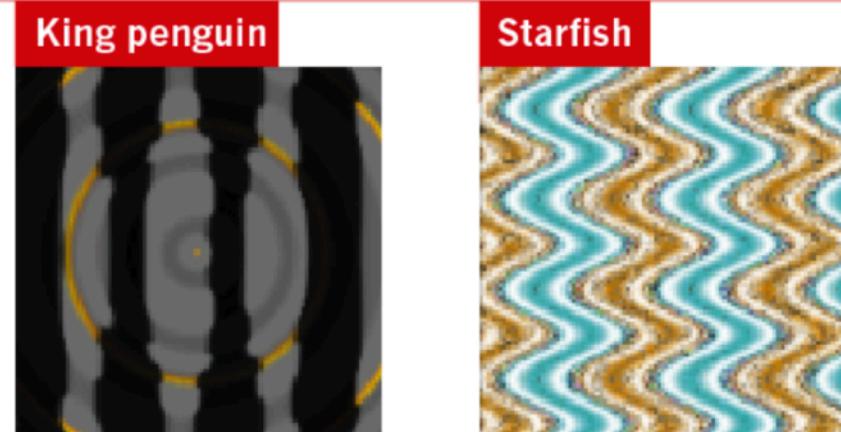
FOOLING THE AI

Deep neural networks (DNNs) are brilliant at image recognition — but they can be easily hacked.

These stickers made an artificial-intelligence system read this stop sign as 'speed limit 45'.



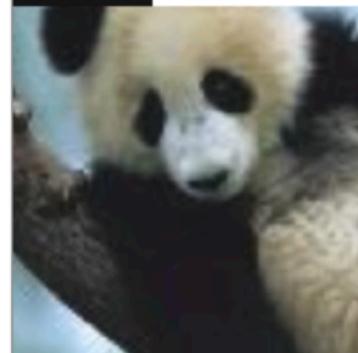
Scientists have evolved images that look like abstract patterns — but which DNNs see as familiar objects.



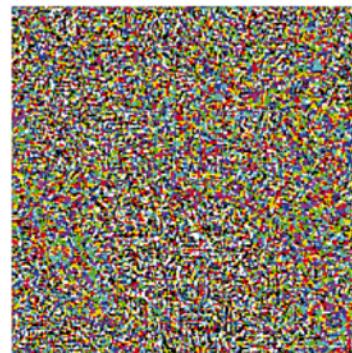
PERCEPTION PROBLEMS

Adding carefully crafted noise to a picture can create a new image that people would see as identical, but which a DNN sees as utterly different.

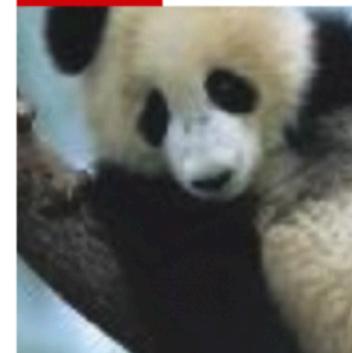
Panda



+



Gibbon



In this way, any starting image can be tweaked so a DNN misclassifies it as any target image a researcher chooses.

Sloth



+

Target image: race car

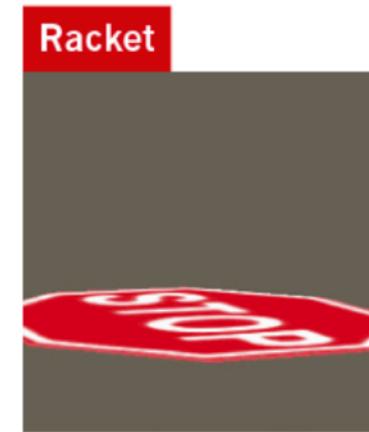
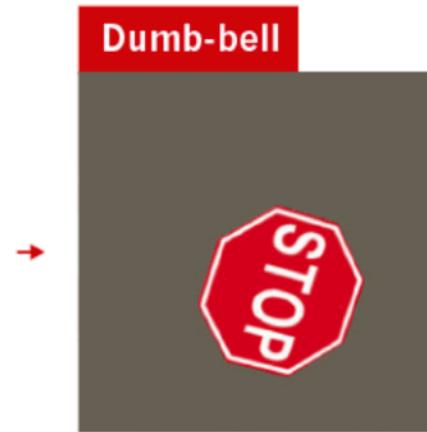


Race car

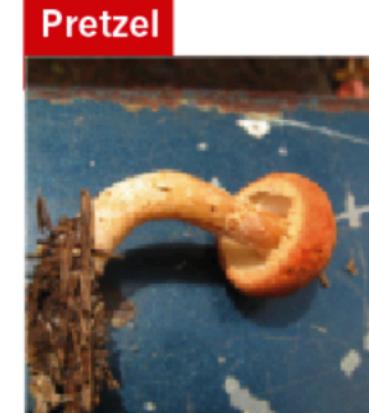


LATEST TRICKS

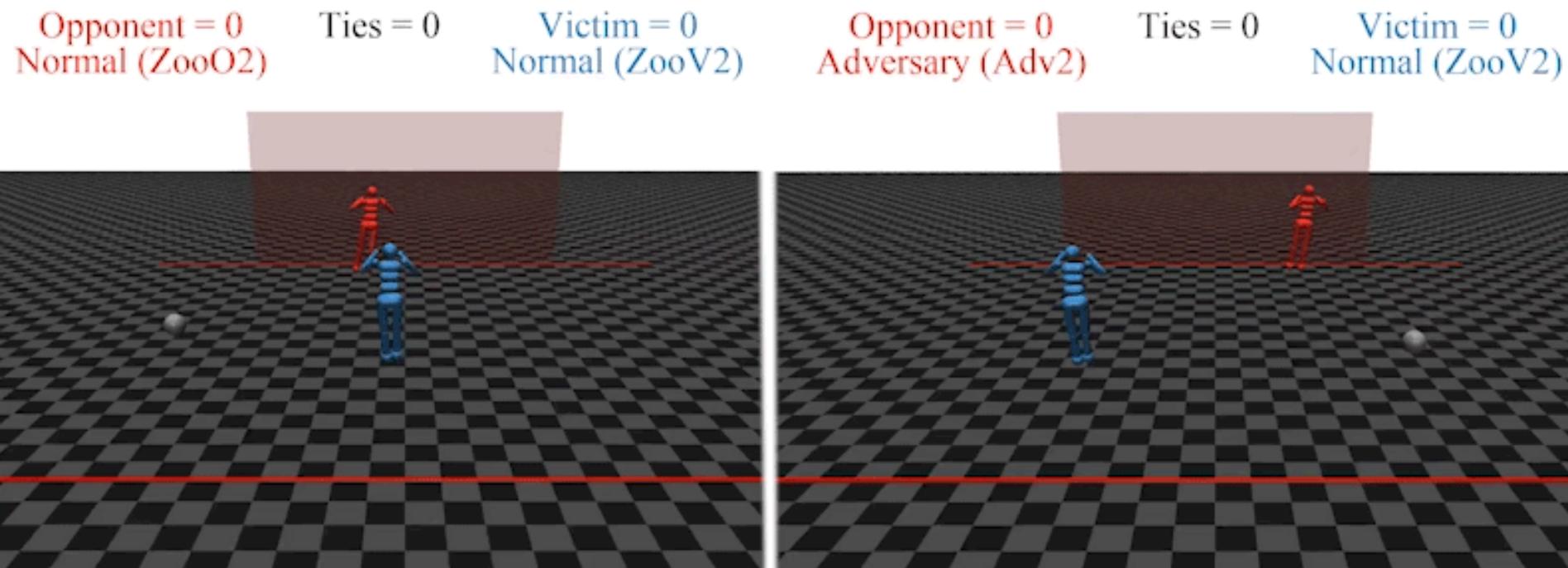
Rotating objects in an image confuses DNNs, probably because they are too different from the types of image used to train the network.



Even natural images can fool a DNN, because it might focus on the picture's colour, texture or background rather than picking out the salient features a human would recognize.

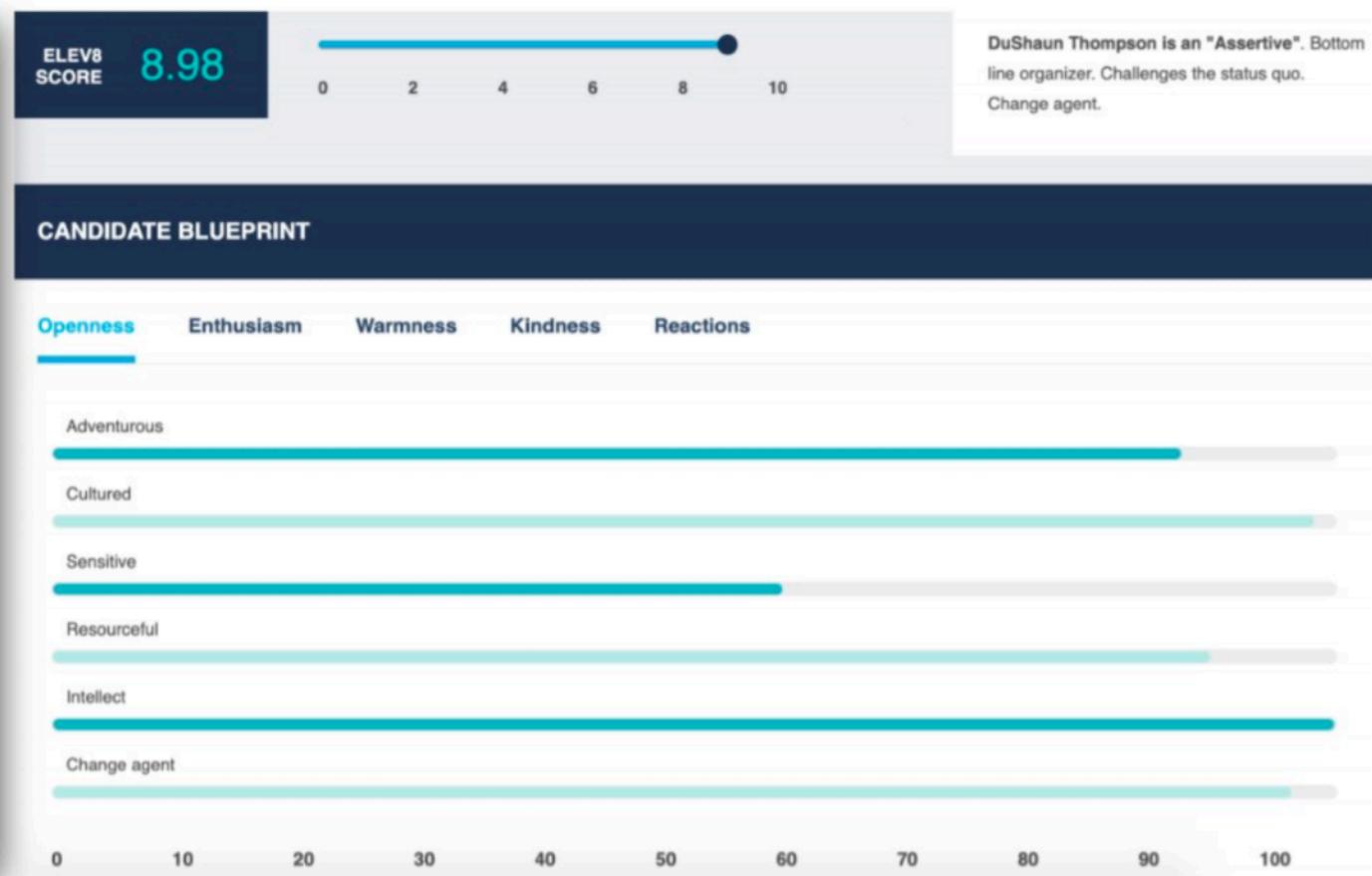


pattern cannot be extrapolated beyond data



source: [Why deep-learning AIs are so easy to fool](#)

Are there really patterns here?



Common sense tells you this isn't possible, and AI experts would agree.
This product is essentially an elaborate random number generator.
- Arvind Narayanan, Assc. Prof. Computer Science @ Princeton

Perception	Judgement	Social Outcome
Image classification Face recognition	Copyright Violation Hate Speech Detection Essay Grading	Job Success Terrorist Risk
Rapid progress	imperfect	dubious



Semakin ambigu bagi manusia
Semakin tidak reliable ML nya

So, beware of how AI is being promoted in media The reality is not as sweet

MIT Technology Review

Artificial intelligence / Machine learning

Google's medical AI was super accurate in a lab. Real life was a different story.

If AI is really going to make a difference to patients we need to know how it works when real humans get their hands on it, in real situations.

by **Will Douglas Heaven**

April 27, 2020

n.b. this is fresh news and it shows how our AI are doing right now (2020)

Others are being predatory too

This AI startup claims to automate app making but actually just uses humans

Who could have seen that coming?

By Nick Statt | @nickstatt | Aug 14, 2019, 1:58pm EDT

The concept of
Mechanical turk



Hype is harmful for business !

but Wira, Is that true?

Most Data Science & AI Project Fails

Why do 87% of data science projects never make it into production?

VB STAFF JULY 19, 2019 4:10 AM

Why 85% of the Artificial Intelligence Projects Fail?

Krunal Vyas - July 11, 2019

WHY 85% OF BIG DATA PROJECTS FAIL

By Sharala Axryd April 16, 2019

- *People from the top must define clear problem statements*
- *Deciding to become data driven can be a long, difficult process*

"One of the biggest opportunities for all of us today is to figure out how we educate the business leaders across the organization," Leff said. "Before, a leader didn't need to necessarily know what the data scientist was doing. Now, the data scientist has stepped into the forefront, and it's actually really important that business leaders understand these concepts."

"It's broken because nobody owned it, we didn't have the data science team to be able to continually iterate on the models, think of it as an asset, and have data operations making sure it's working well," Chapo said. "We're starting to

"I've had data scientists look me in the face and say we could do that project, but we can't get access to the data," Leff says. "And I say, your management allows that to go on?"

Summary of Most common problem

- Wrong questions
 - Not clearly defined
 - No clear lead to business value
- Lack of support
 - Commitment to implement changes
 - No proper strategy
- Data Problems
 - Data availability
 - Poor data quality
- Incomplete Data Science Team
 - Business Internals, ETL, EDA, Modeling, Deployment
- Over promising
 - Solution does not appear right away
 - Slow and small refinement

n.b. Red colors are related to overhype and over expectation

Why so pessimistic? Any Solution?

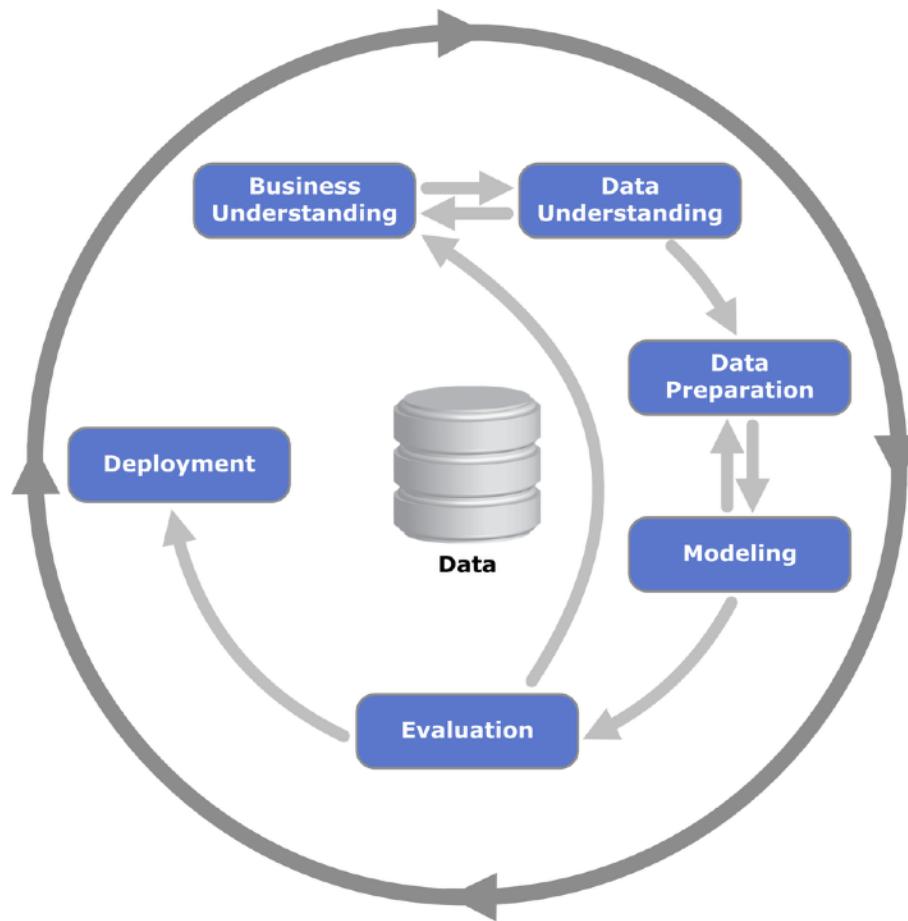
We are progressively improving, but for now we can use

A Proper Process in Data Science

9 Popular Project Management Methodologies

- 1  Agile – collaborating to iteratively deliver whatever works
- 2  Scrum – enabling a small, cross-functional, self-managing team to deliver fast
- 3  Kanban – improving speed and quality of delivery by increasing visibility of work in progress and limiting multi-tasking
- 4  Scrumban – limiting work in progress like Kanban, with a daily stand up like Scrum
- 5  Lean – streamlining and eliminating waste to deliver more with less
- 6  eXtreme Programming (XP) – doing development robustly to ensure quality
- 7  Waterfall – planning projects fully, then executing through phases
- 8  PRINCE2 – controlled project management that leaves nothing to chance
- 9  PMI's PMBOK – applying universal standards to Waterfall project management

Agile does not work pretty well in data science



CRISP-DM

Summary on Data Science implementation

=> Definisikan masalah / kebutuhan bisnis yang ingin dijawab

=> bangun Tim Business Intelligence (BI)

- analisa **data masa lampau**
- mostly **structured / tabular data**

=> bangun Tim Business Analytics (BA)

- **modeling dan prediksi**
- **menganalisa unstructured data**

Closing Remarks

- I have shared with you the **sweet hype** and the **sour truth**
- Be objective and **start learning with proper mindset.**
- This is the General Intro, still **lots of details to learn**
- There are **various job roles**, each has their own tasks
- **Not all problems need Machine Learning**
- Deep Learning is used mostly on unstructured data. **You may not even need it.**

References

- [Raj Ramesh - What are the different roles in Data Science?](#)
- [AI, Deep Learning, and Machine Learning: A Primer](#)
- [Don't fear intelligent machines. Work with them by Garry Kasparov](#)
- [Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm](#)
- [Snake AI Fight](#)
- [Pneumonia classification with PyTorch](#)
- [Meet the 26-Year-Old Hacker Who Built a Self-Driving Car... in His Garage](#)
- [AllenNLP Named Entity Recognition](#)
- [Huggingface Transformer Demo](#)
- [Hierarchical Surface Prediction for 3D Object Reconstruction](#)
- [Photorealistic Facial Texture Inference Using Deep Neural Networks](#)
- [Visually-indicated sounds](#)
- [Text to Speech Bahasa Indonesia menggunakan Deep Learning](#)
- [MIDI Music Composition with Deep Learning – Code in Python](#)
- [Neural Style Transfer](#)
- [Bhinneka - Hendrik Tio Face Recognition](#)
- [Faceswap with DFL Styled Autoencoder](#)
- [CycleGAN](#)
- [Amazon doesn't care if you accidentally shoplift from its cashier-less store](#)
- [Why deep-learning AIs are so easy to fool](#)
- [How to recognize AI Snake Oil](#)
- [Google's medical AI was super accurate in a lab. Real life was a different story.](#)
- [The Turk](#)
- [87% Data Science projects never make into production](#)
- [6 reason data science projects fail](#)
- [9 of the most popular Project Management Methodologies](#)
- [CRISP-DM](#)
- [Image Icon, Document Icon, Spreadsheet Icon](#)
- [Logo Peringatan HUT RI ke-75](#)

That's it for now
Thank you for your attention

Wira Dharma Kencana Putra



/in/wiradkputra



[youtube.com/jcopuntukindonesia](https://www.youtube.com/jcopuntukindonesia)



wiradharma_kencanaputra@yahoo.com

Please add **[JCOp]** to your subject if you would like to contact me by email