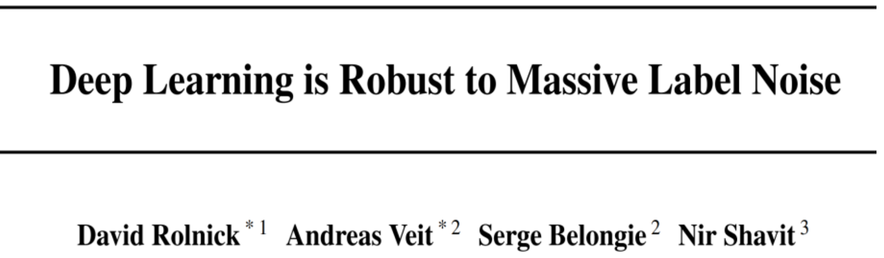
Distributed Algorithms as a Gateway to Thinking Slow

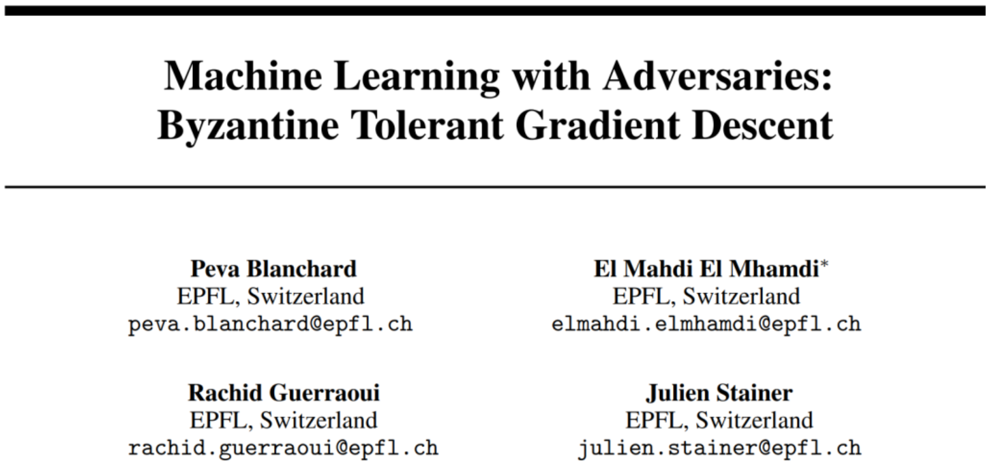
Roger Wattenhofer

ML Party



Deep Learning is Robust to Massive Label Noise

David Rolnick [*1] Andreas Veit [*2] Serge Belongie [2] Nir Shavit [3]



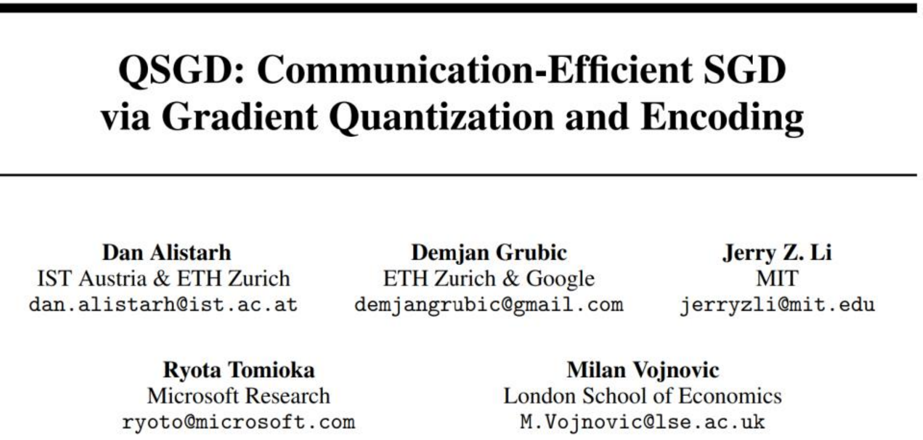
Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent

Peva Blanchard
EPFL, Switzerland
`peva.blanchard@epfl.ch`

El Mahdi El Mhamdi[*]
EPFL, Switzerland
`elmahdi.elmhamdi@epfl.ch`

Rachid Guerraoui
EPFL, Switzerland
`rachid.guerraoui@epfl.ch`

Julien Stainer
EPFL, Switzerland
`julien.stainer@epfl.ch`



QSGD: Communication-Efficient SGD via Gradient Quantization and Encoding

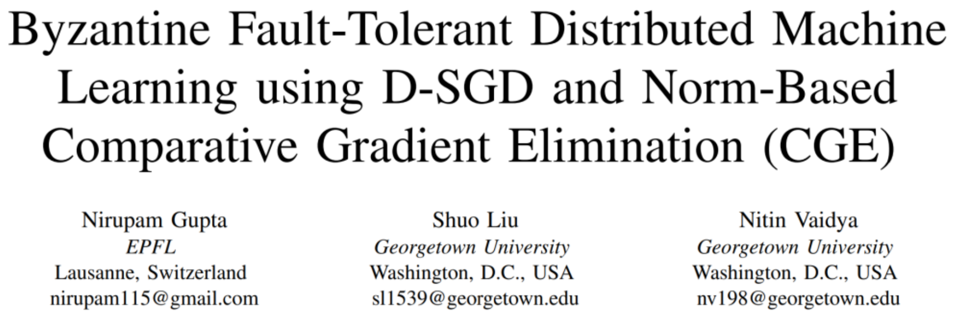
Dan Alistarh
IST Austria & ETH Zurich
`dan.alistarh@ist.ac.at`

Demjan Grubic
ETH Zurich & Google
`demjangrubic@gmail.com`

Jerry Z. Li
MIT
`jerryzli@mit.edu`

Ryota Tomioka
Microsoft Research
`ryoto@microsoft.com`

Milan Vojnovic
London School of Economics
`M.Vojnovic@lse.ac.uk`



Byzantine Fault-Tolerant Distributed Machine Learning using D-SGD and Norm-Based Comparative Gradient Elimination (CGE)

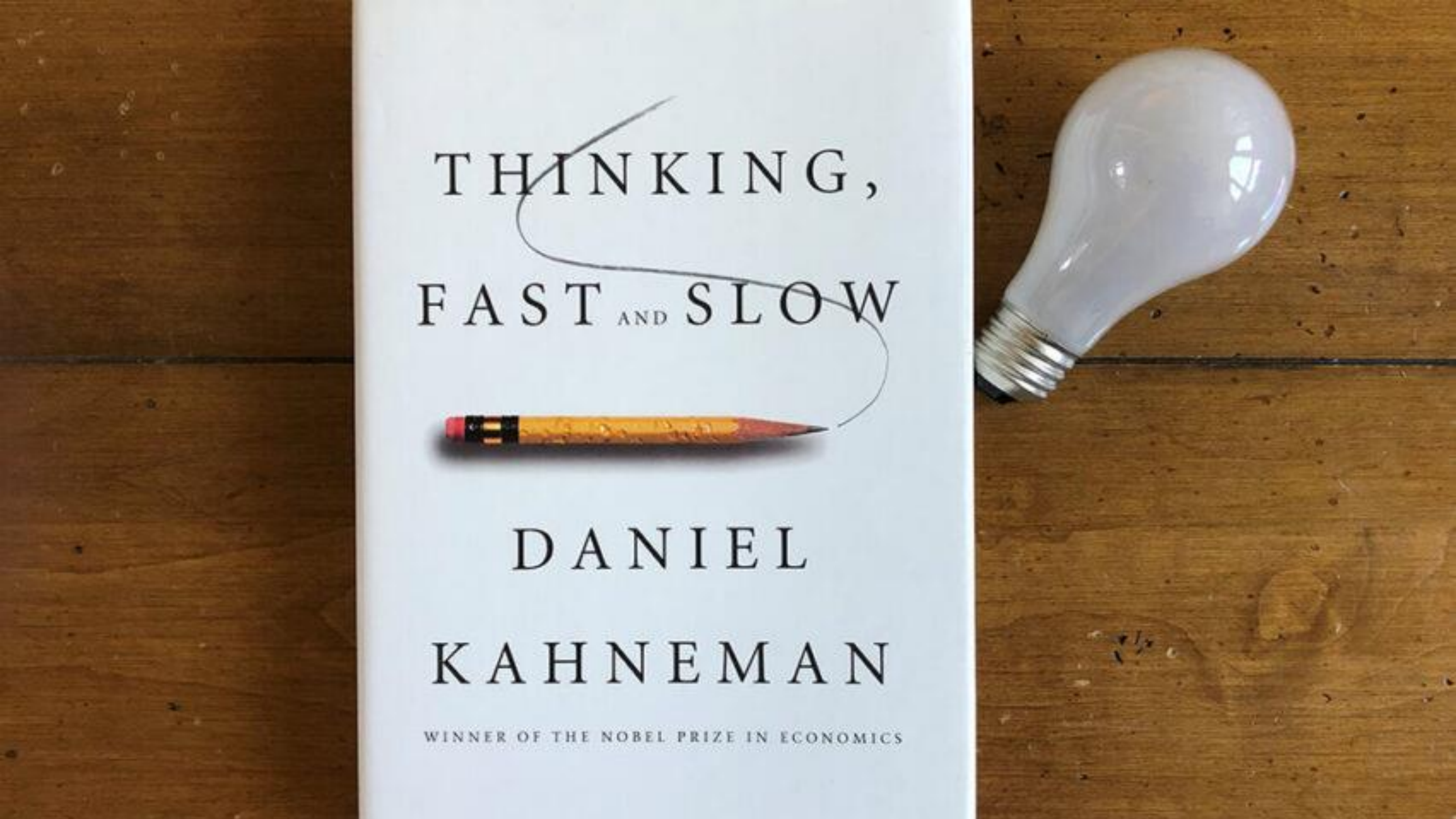
Nirupam Gupta
EPFL
Lausanne, Switzerland
nirupam115@gmail.com

Shuo Liu
Georgetown University
Washington, D.C., USA
sl1539@georgetown.edu

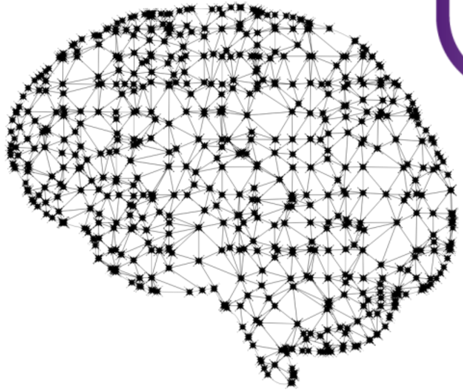
Nitin Vaidya
Georgetown University
Washington, D.C., USA
nv198@georgetown.edu

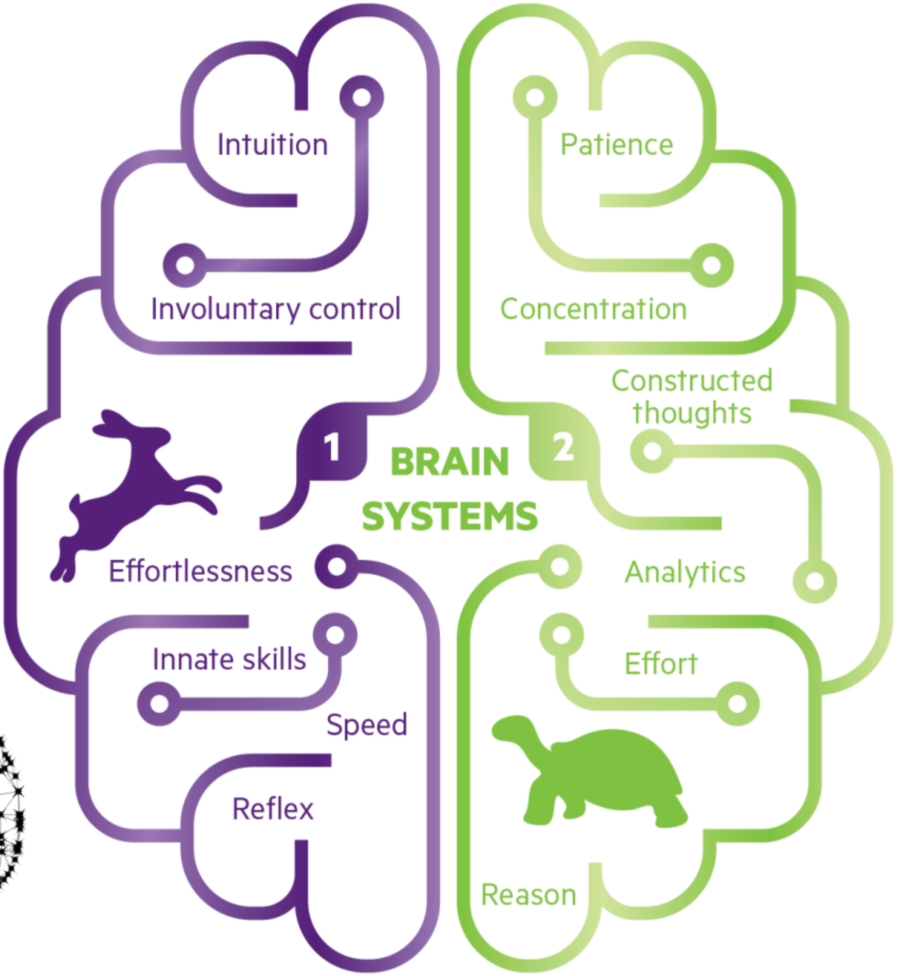
Distributed Algorithms as a Gateway to Thinking Slow

Roger Wattenhofer


THINKING,
FAST AND SLOW
DANIEL
KAHNEMAN
WINNER OF THE NOBEL PRIZE IN ECONOMICS

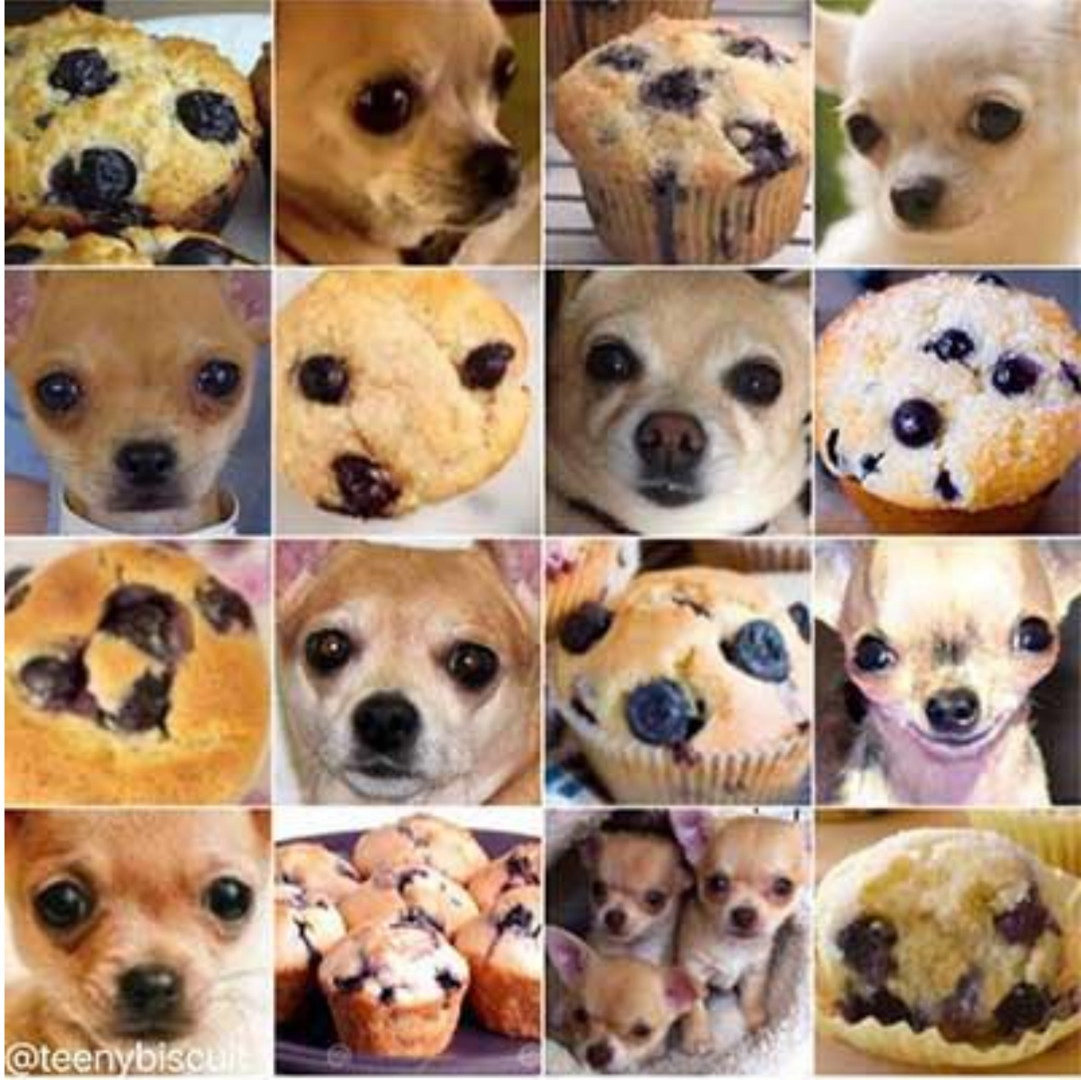
Machine Learning



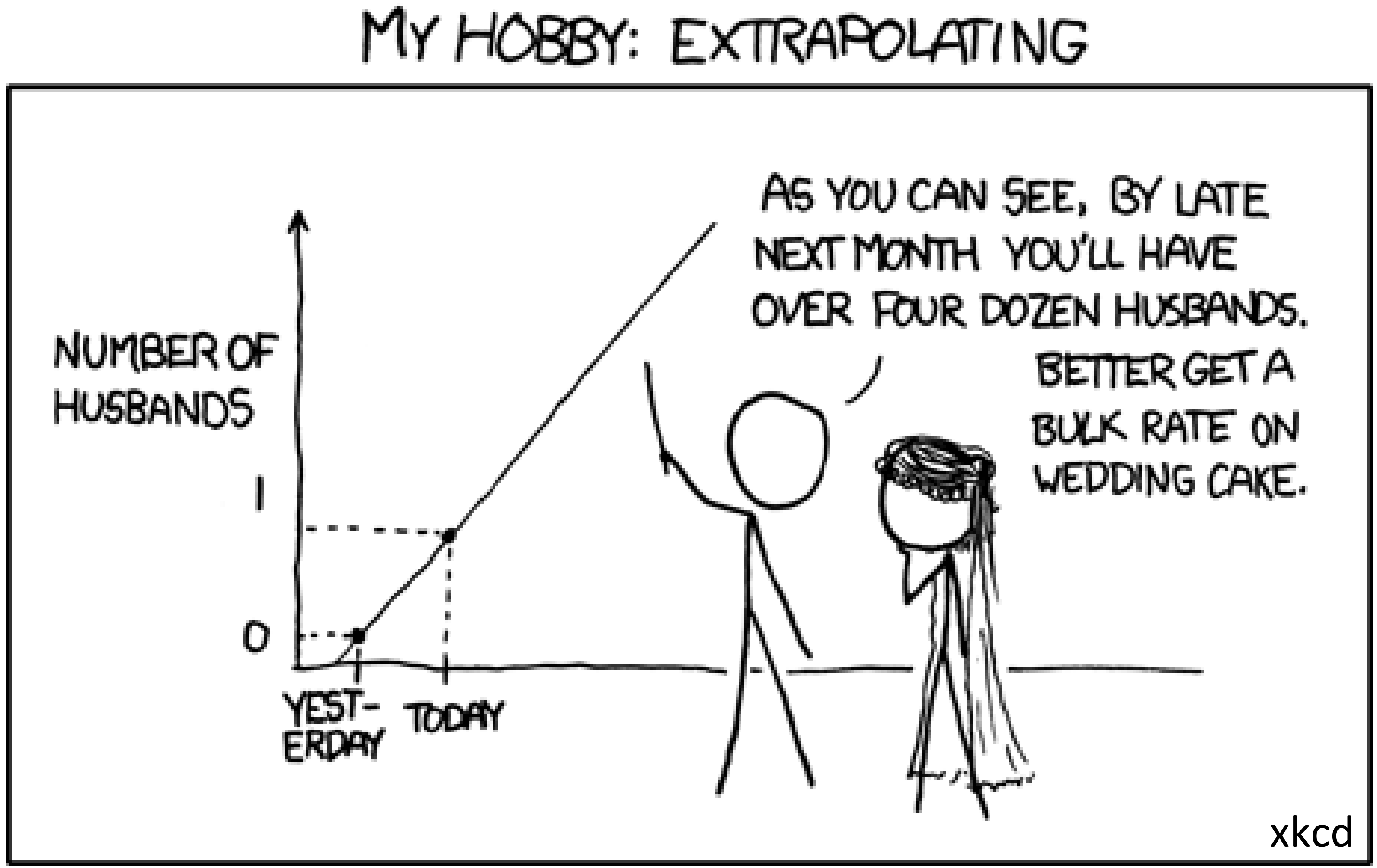


Classification

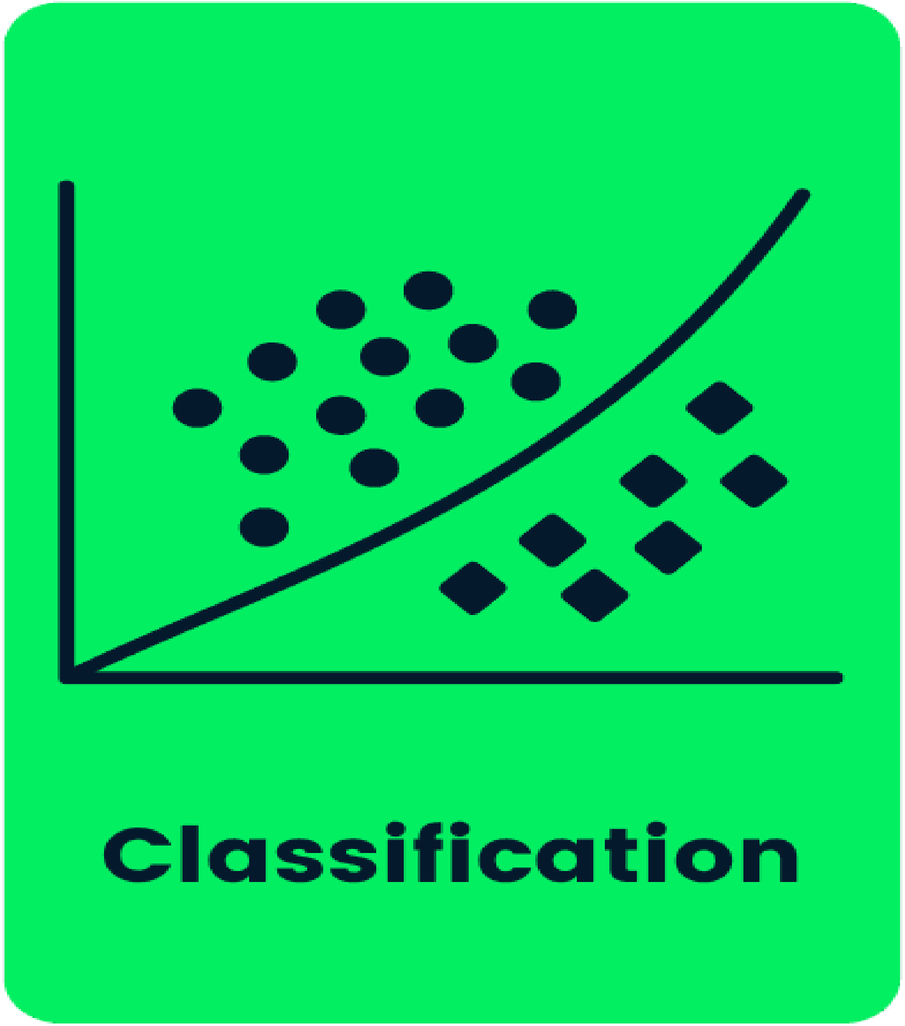
Chihuahua
vs.
Muffin

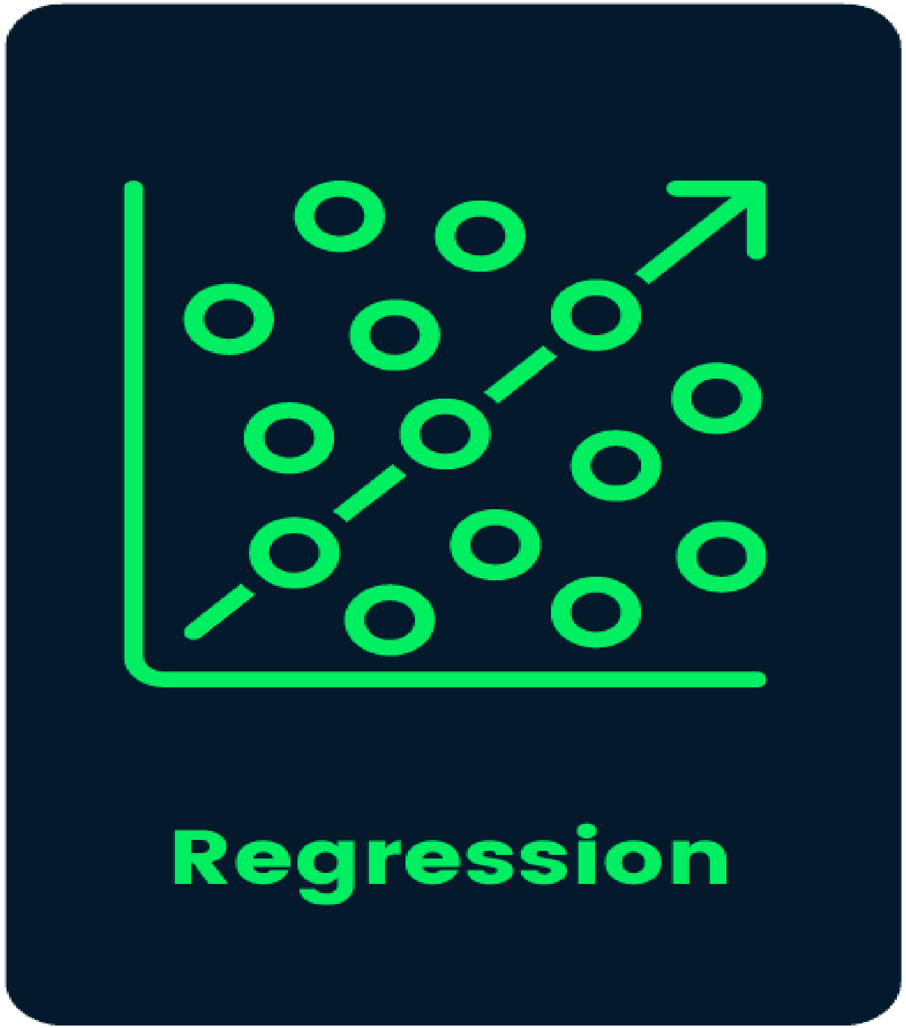


Linear Regression


MY HOBBY: EXTRAPOLATING
NUMBER OF HUSBANDS
1
0
YEST-ERDAY
TODAY
AS YOU CAN SEE, BY LATE NEXT MONTH YOU'LL HAVE OVER FOUR DOZEN HUSBANDS. BETTER GET A BULK RATE ON WEDDING CAKE.

xkcd


Classification


Regression

Classification

Regression

Forecast

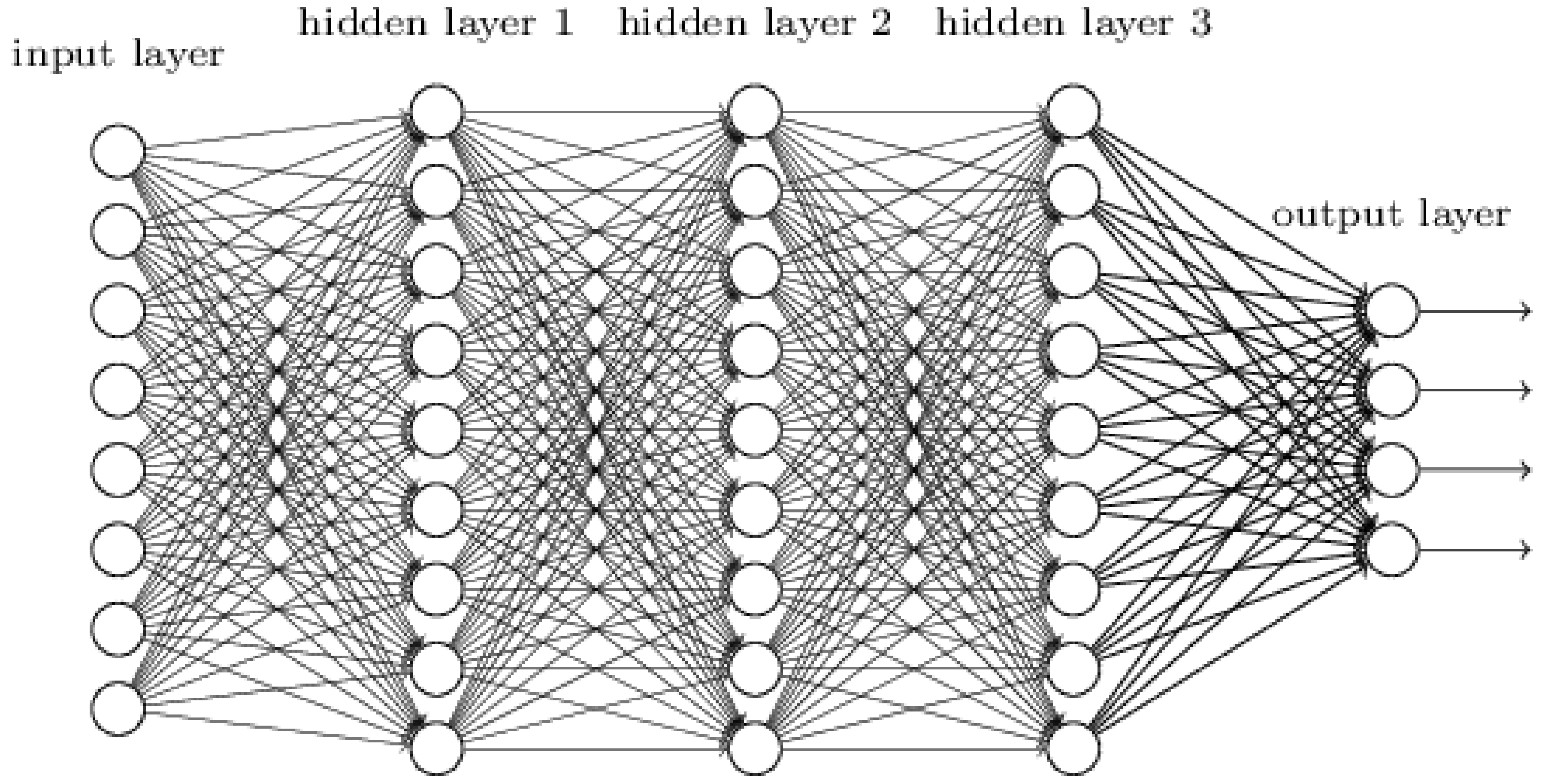
Clustering

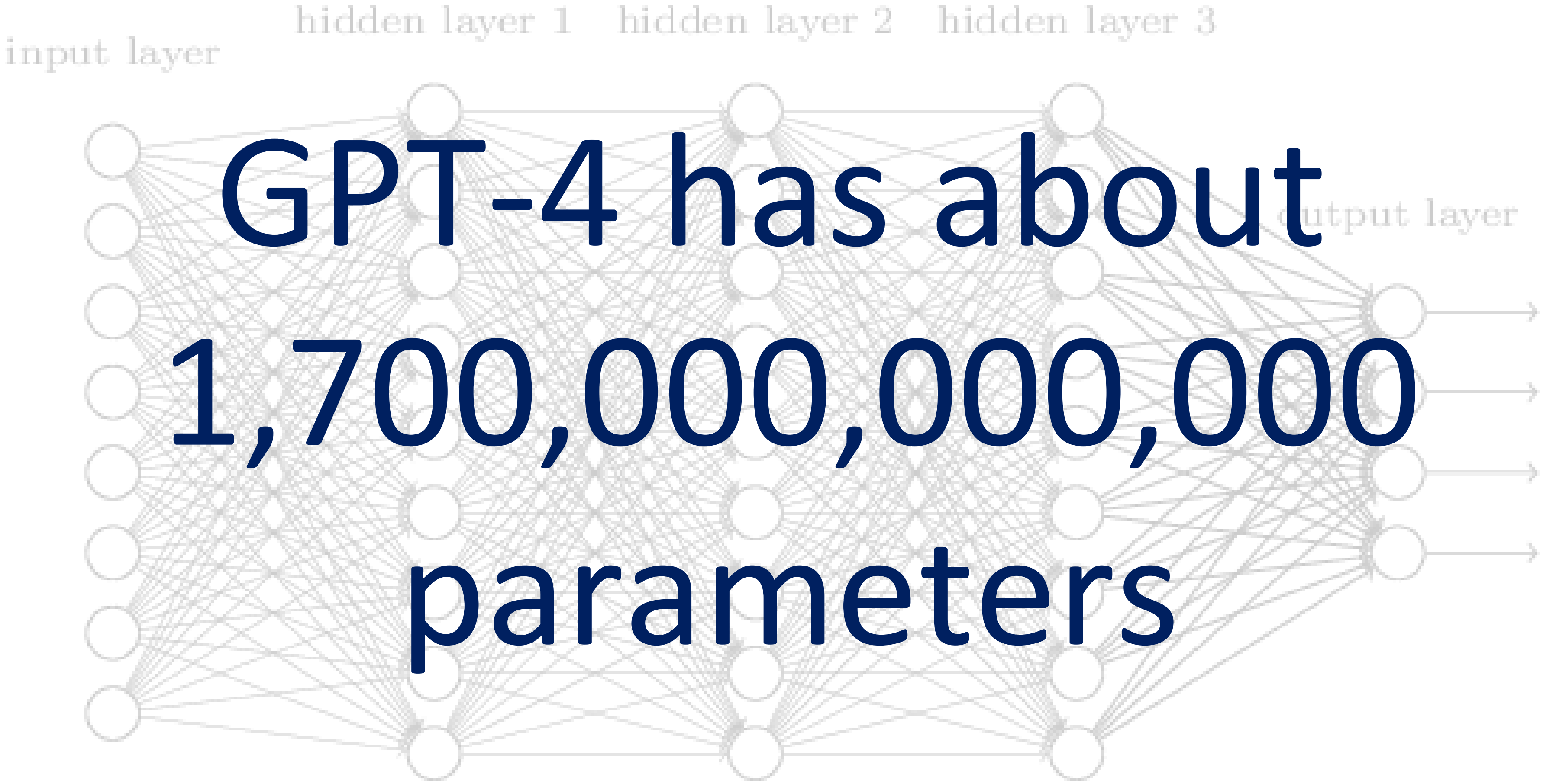
Machine Translation

Computer Vision

Generative Art

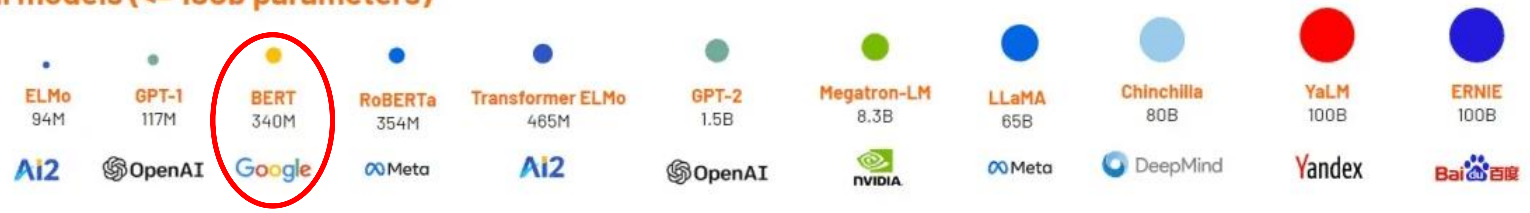
Reinforcement learning


input layer
hidden layer 1
hidden layer 2
hidden layer 3
output layer

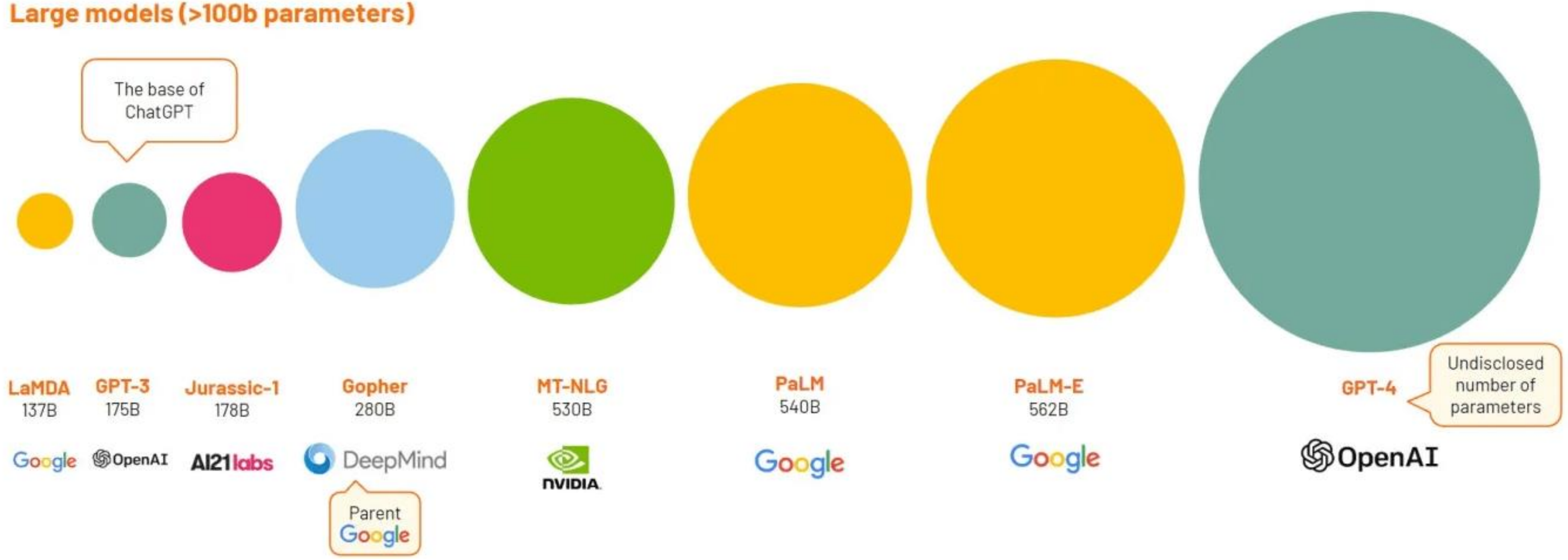


GPT-4 has about 1,700,000,000,000 parameters

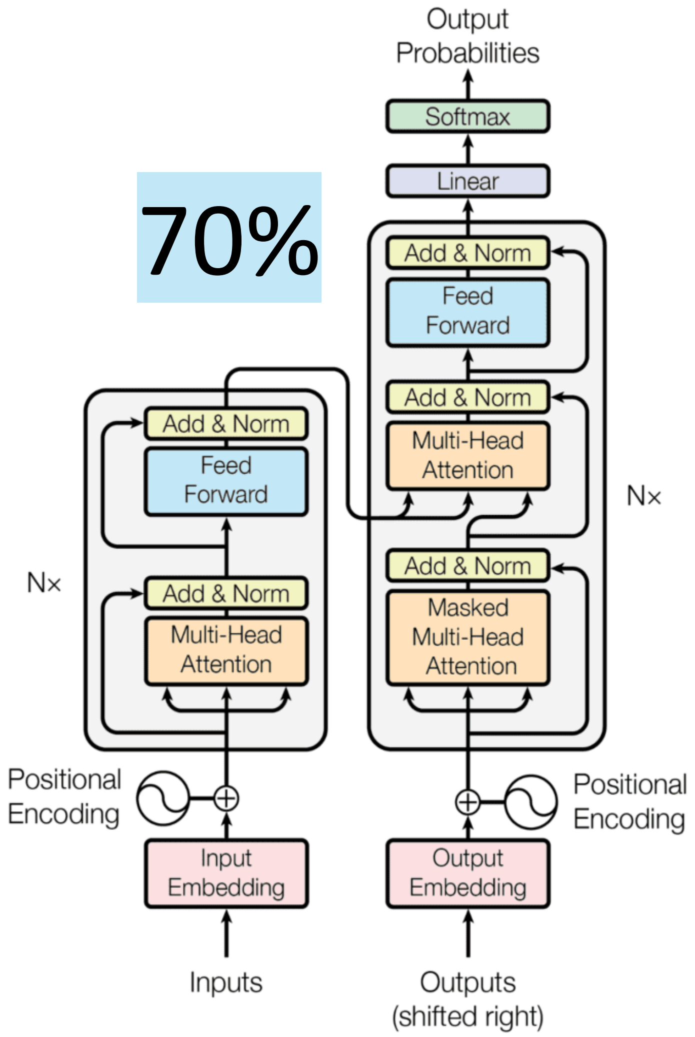
Small models (<= 100b parameters)

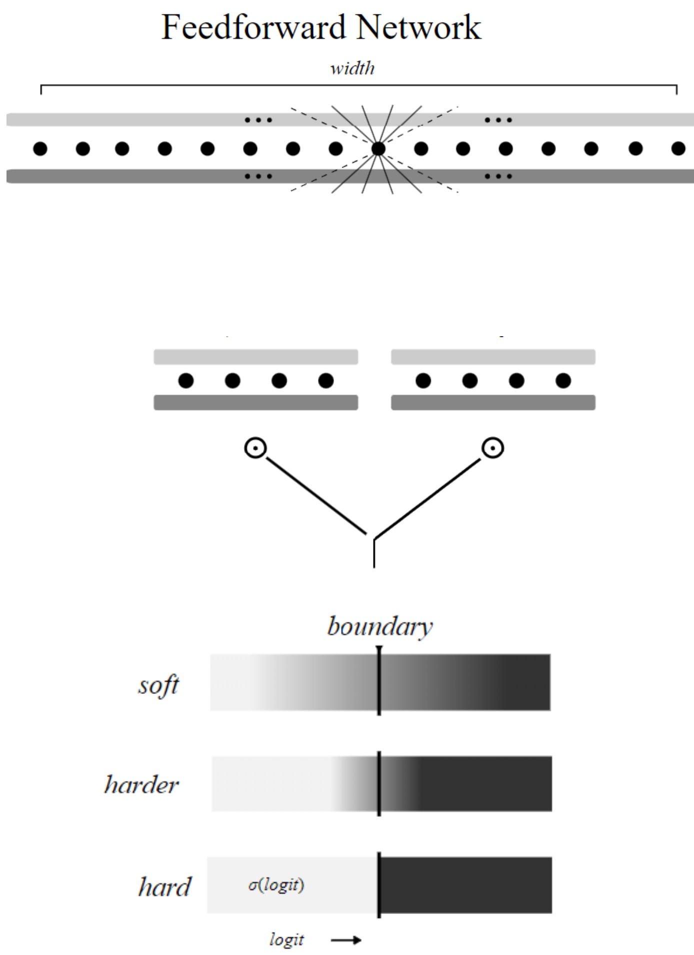


Large models (>100b parameters)



70%

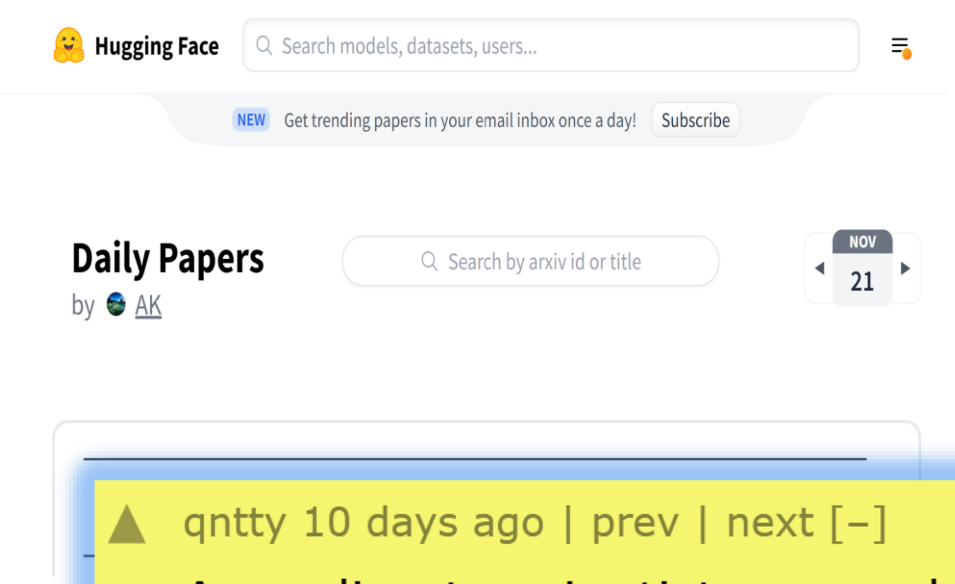

Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)


Feedforward Network
width
boundary
soft
harder
hard
σ(logit)
logit

Exponentially Faster Language Modeling

Peter Belcak and Roger Wattenhofer
ETH Zürich
`{belcak,wattenhofer}@ethz.ch`

Model	N_T	$N_\mathrm{I}/N_\mathrm{T}$	RTE	MRPC	STSB	SST-2	MNLI	QNLI	QQP	Avg	CoLA	Avg
UltraFastBERT-1x11-long	4095	0.3%	60.7	87.5	86.4	89.9	81.3	89.7	87.6	83.0	35.1	77.7
External Baselines												
OpenAI GPT	3072	100%	56.0	82.3	80.0	91.3	81.4	87.4	70.3	78.8	45.4	75.1
DistilBERT	3072	100%	59.9	87.5	86.9	91.3	82.2	89.2	71.3	81.2	52.1	77.6
BERT-base	3072	100%	66.4	88.9	85.8	93.5	83.4	90.5	71.2	83.0	51.3	79.6


Hugging Face
Search models, datasets, users...
NEW Get trending papers in your email inbox once a day! Subscribe
Search by arxiv id or title
NOV
21

Daily Papers

by AK

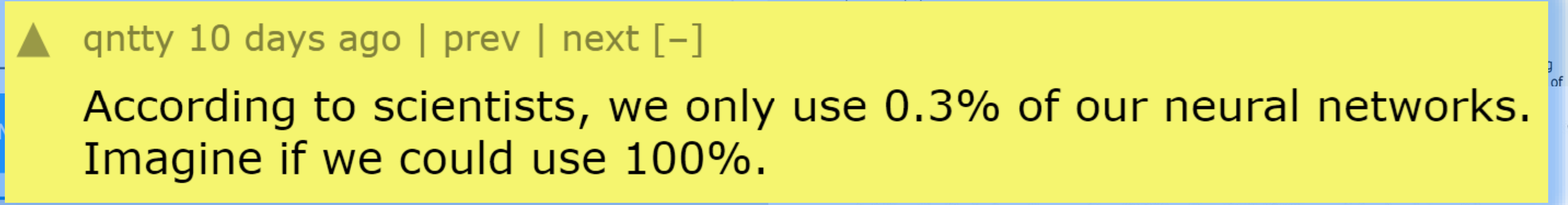
Hacker News new | past | comments | ask | show | jobs | submit
login
add comment

Exponentially faster language modelling (arxiv.org)

300 points by born-jre 10 days ago | hide | past | favorite | 137 comments

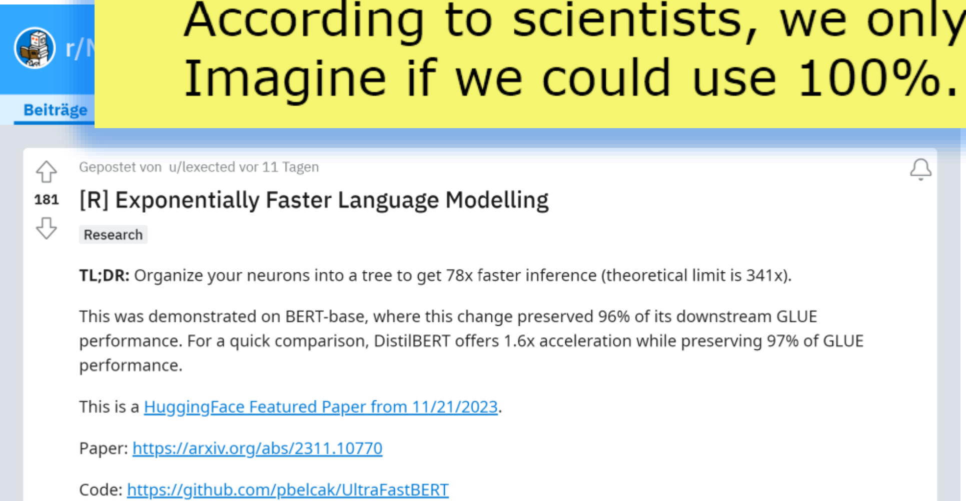
WithinReason 10 days ago | next [–]

Link to previous paper:



qntty 10 days ago | prev | next [–]

According to scientists, we only use 0.3% of our neural networks. Imagine if we could use 100%.



Gepostet von u/lexected vor 11 Tagen

181

[R] Exponentially Faster Language Modelling

Research

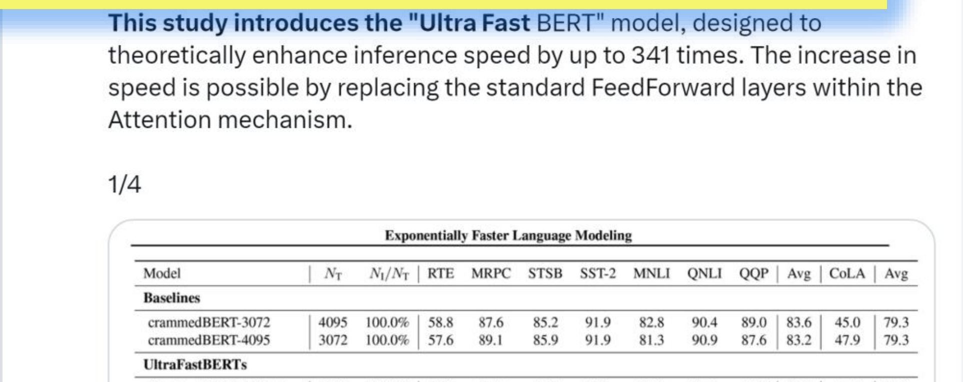
TL;DR: Organize your neurons into a tree to get 78x faster inference (theoretical limit is 341x).

This was demonstrated on BERT-base, where this change preserved 96% of its downstream GLUE performance. For a quick comparison, DistilBERT offers 1.6x acceleration while preserving 97% of GLUE performance.

This is a HuggingFace Featured Paper from 11/21/2023.

Paper: https://arxiv.org/abs/2311.10770

Code: https://github.com/pbelcak/UltraFastBERT

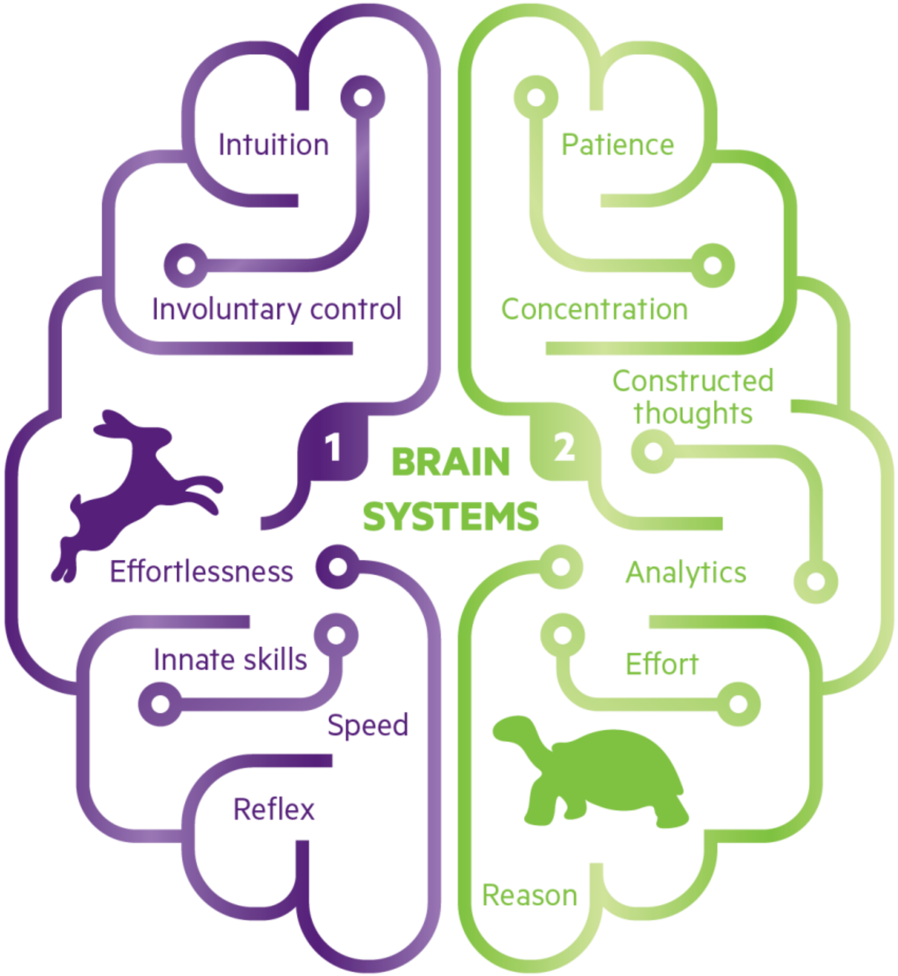


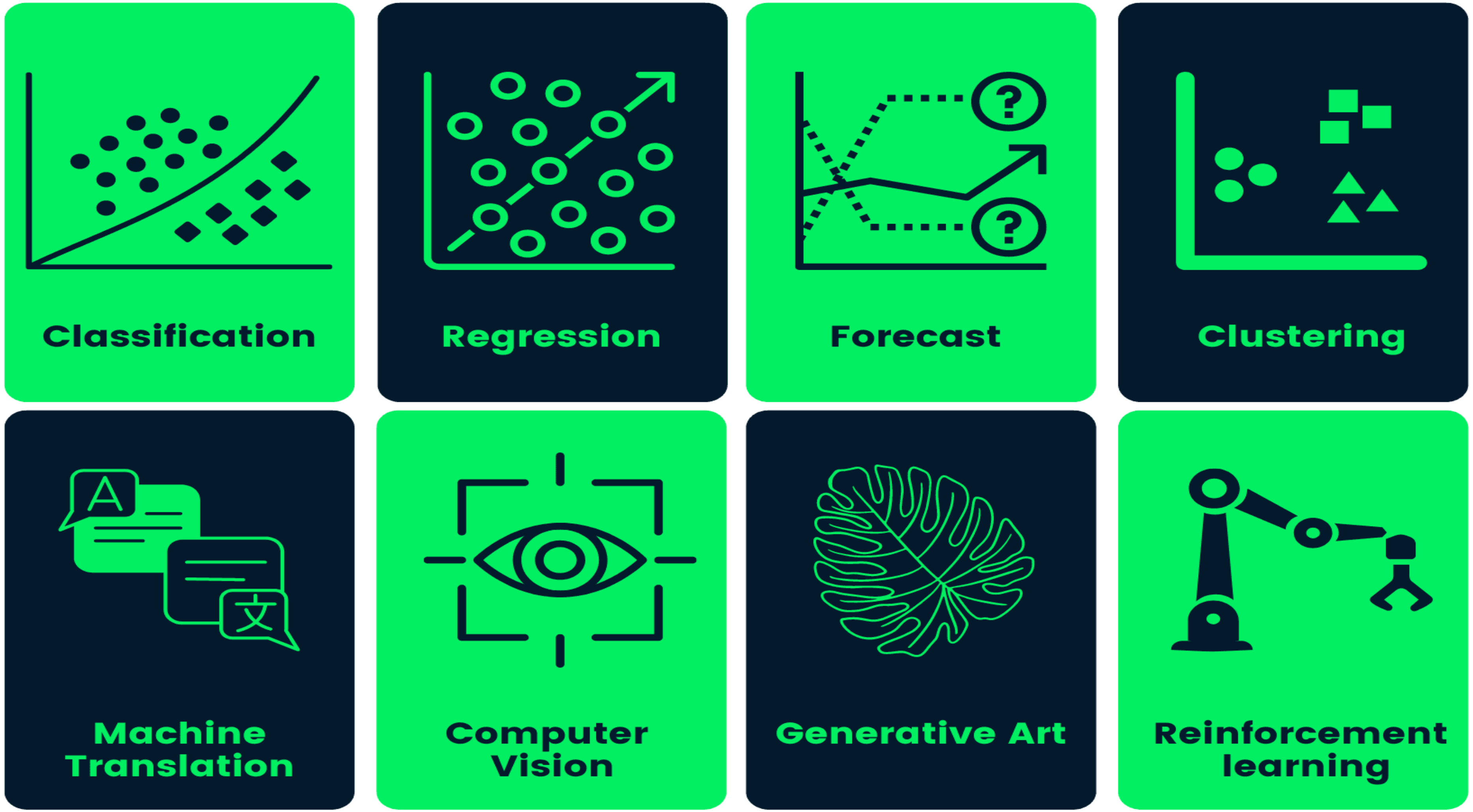
This study introduces the "Ultra Fast BERT" model, designed to theoretically enhance inference speed by up to 341 times. The increase in speed is possible by replacing the standard FeedForward layers within the Attention mechanism.

1/4

Exponentially Faster Language Modeling

Model	N_T	N_I/N_T	RTE	MRPC	STSB	SST-2	MNLI	QNLI	QQP	Avg	CoLA	Avg
Baselines												
crammedBERT-3072	4095	100.0%	58.8	87.6	85.2	91.9	82.8	90.4	89.0	83.6	45.0	79.3
crammedBERT-4095	3072	100.0%	57.6	89.1	85.9	91.9	81.3	90.9	87.6	83.2	47.9	79.3
UltraFastBERTs												

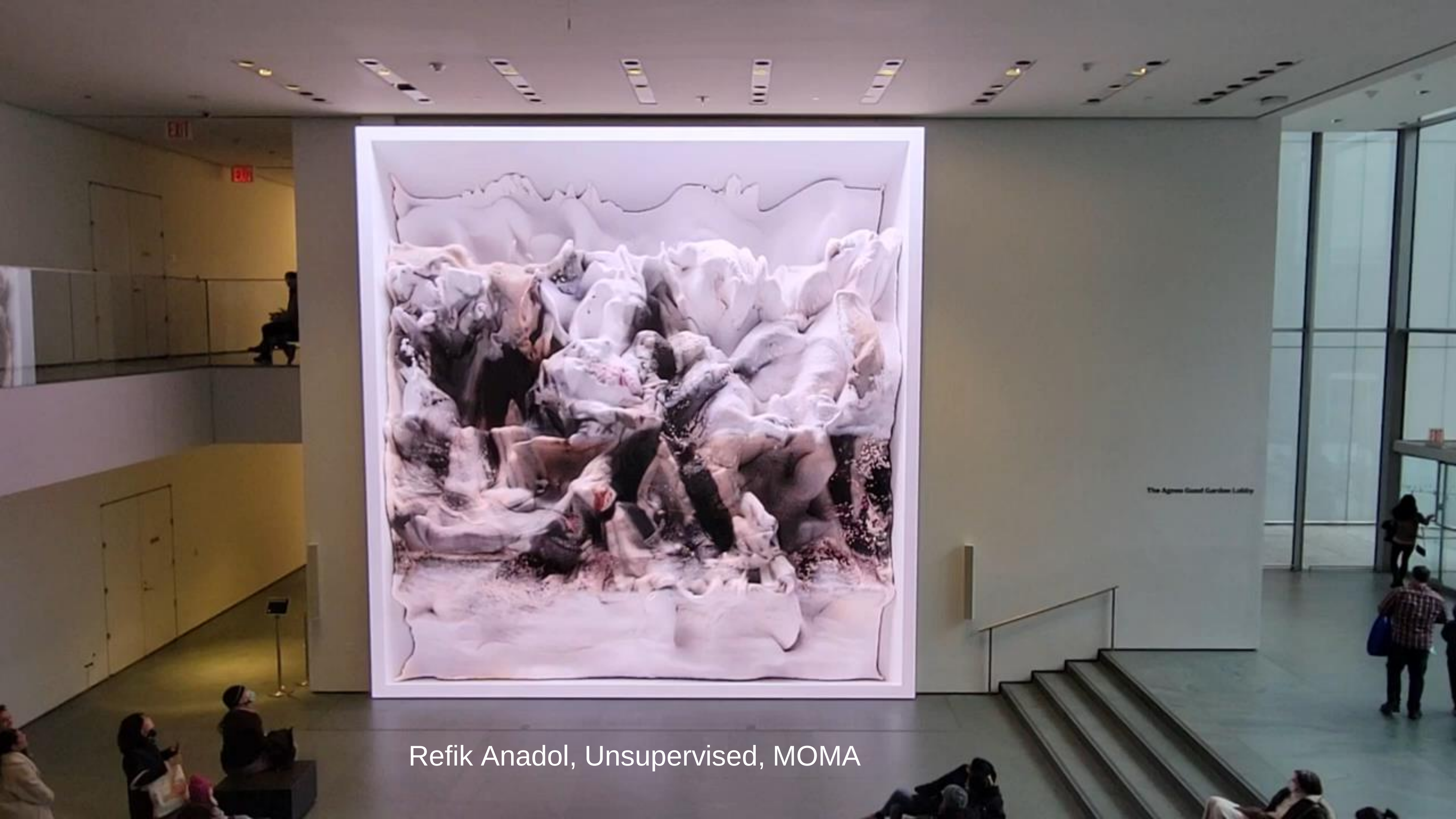

Intuition
Involuntary control
Effortlessness
Innate skills
Speed
Reflex
1
BRAIN SYSTEMS
2
Patience
Concentration
Constructed thoughts
Analytics
Effort
Reason


Classification
Regression
Forecast
Clustering
Machine Translation
Computer Vision
Generative Art
Reinforcement learning

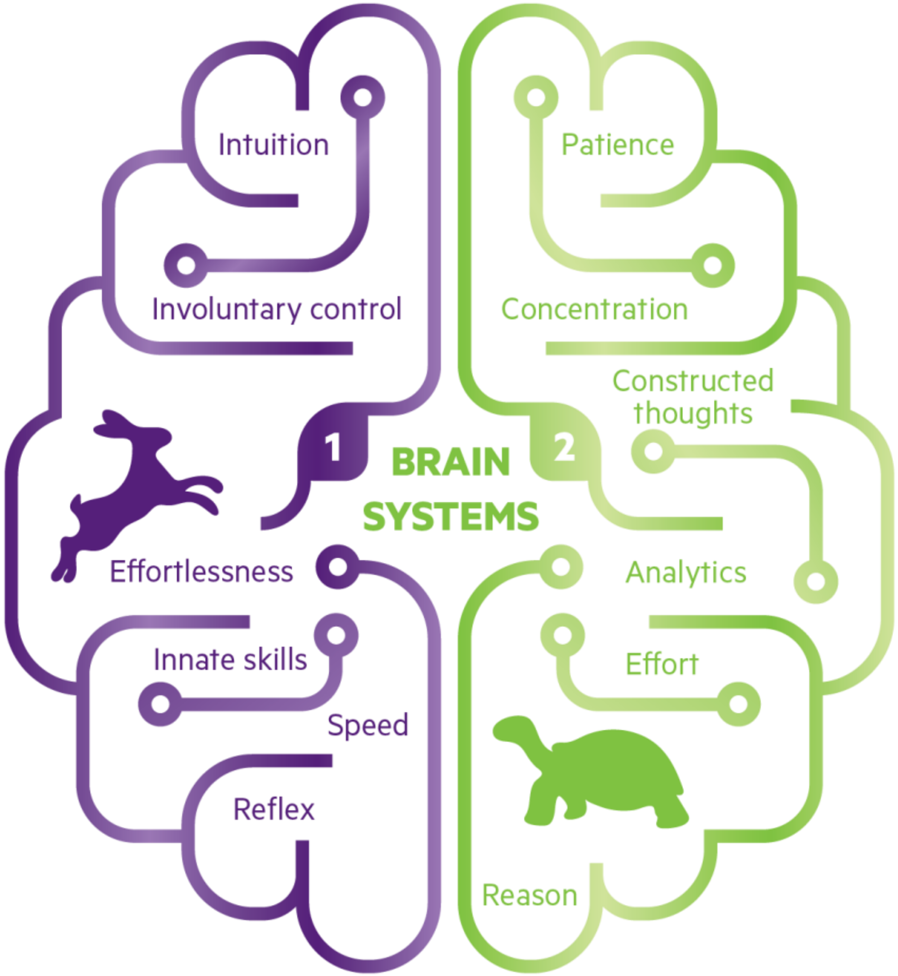
uter
on

Generative Art

Reinfor
lear



Refik Anadol, Unsupervised, MOMA


Intuition
Involuntary control
Effortlessness
Innate skills
Speed
Reflex
1
BRAIN SYSTEMS
2
Patience
Concentration
Constructed thoughts
Analytics
Effort
Reason

User

what is 5 times 7?

Assistant

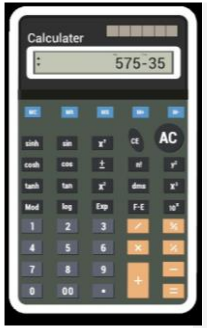
5 times 7 equals 35.

User

what is 384903720 times 538982734?

Assistant

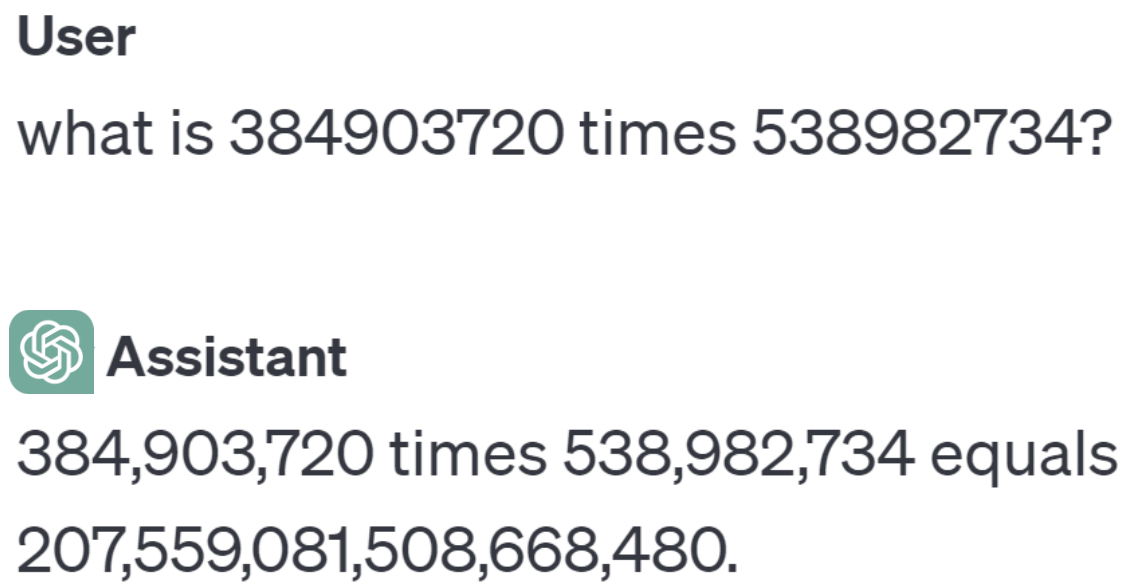
384,903,720 times 538,982,734 equals 207,559,081,508,668,480.



384903720 × 538982734 =

207,456,459,332,370,480

$$\begin{array}{r}
314 \\
159\ \times \\
\hline
{}^{1}2\,8\,{}^{1}2\,{}^{3}6 \\
1\,5\,7\,{}^{2}0\,0 \\
3\,1\,4\,0\,0 \\
\hline
4\,9\,9\,2\,6
\end{array}$$



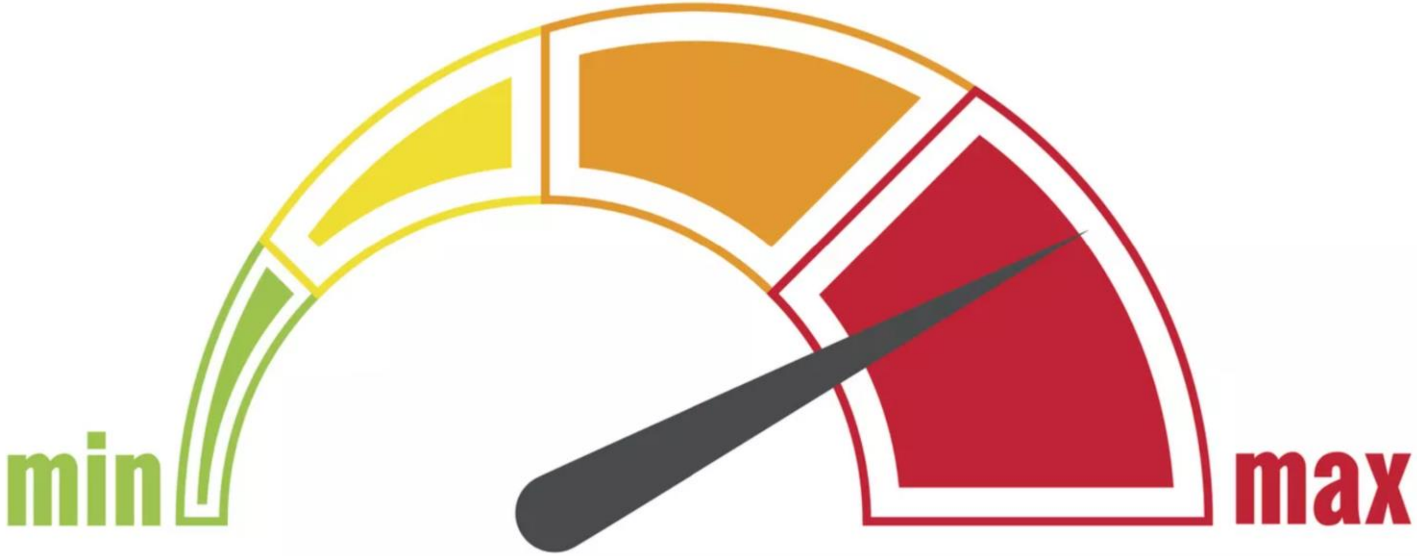
User

what is 384903720 times 538982734?

Assistant

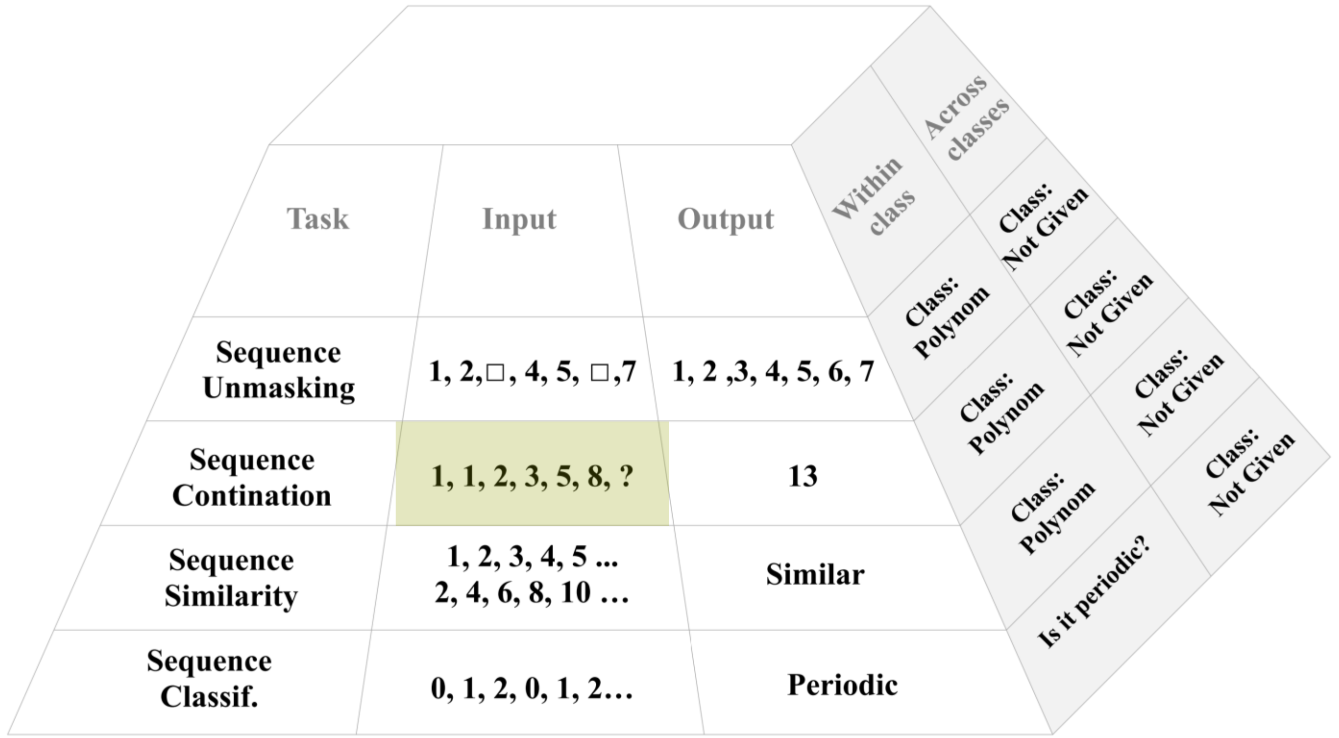
384,903,720 times 538,982,734 equals 207,559,081,508,668,480.

Thinking Slow Benchmark?

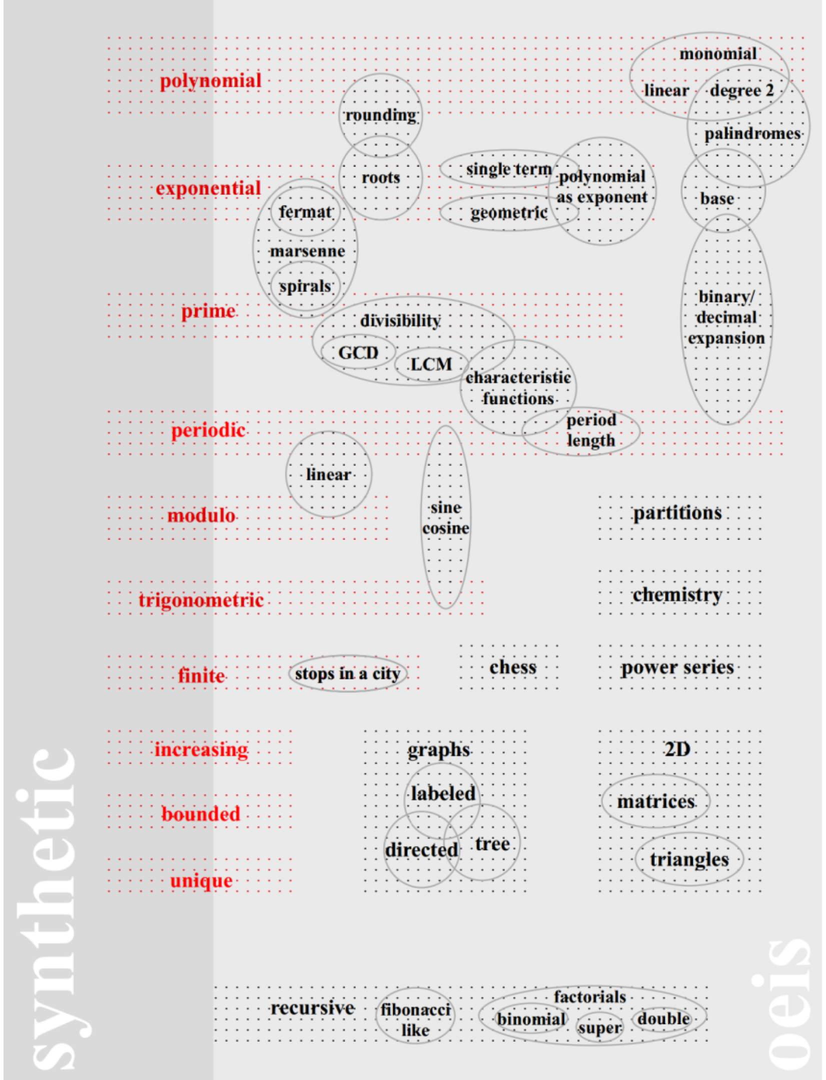


Fact: Learning Governing Abstractions Behind Integer Sequences

Peter Belcák, Ard Kastrati, Flavio Schenker, Roger Wattenhofer

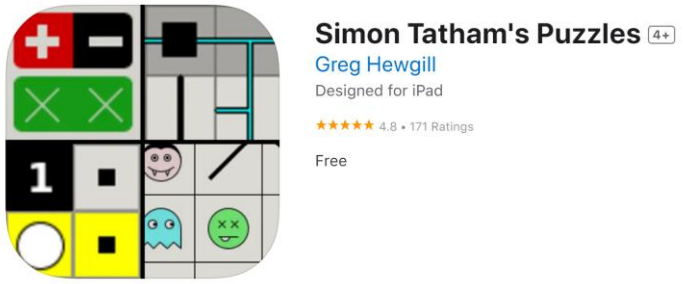


Task	Input	Output	Within class	Across classes
Sequence Unmasking	1, 2,□, 4, 5, □,7	1, 2 ,3, 4, 5, 6, 7	Class: Polynom	Class: Not Given
Sequence Contination	1, 1, 2, 3, 5, 8, ?	13	Class: Polynom	Class: Not Given
Sequence Similarity	1, 2, 3, 4, 5 ... 2, 4, 6, 8, 10 …	Similar	Class: Polynom	Class: Not Given
Sequence Classif.	0, 1, 2, 0, 1, 2…	Periodic	Is it periodic?	Class: Not Given


polynomial
exponential
prime
periodic
modulo
trigonometric
finite
increasing
bounded
unique
synthetic
oeis
monomial
linear
degree 2
palindromes
rounding
roots
single term
geometric
polynomial as exponent
base
fermat
marsenne
spirals
divisibility
GCD
LCM
characteristic functions
period length
binary/ decimal expansion
linear
sine cosine
partitions
chemistry
stops in a city
chess
power series
graphs
labeled
directed
tree
2D
matrices
triangles
recursive
fibonacci like
factorials
binomial
super
double

パズル通信ニコリ
季刊
2022年春号
Vol. 178
特集 遊園地は良いぞ
数独通信
数独好きがつながる雑誌
ニコリの謎解き
ひらめきパズル
with クロスワード 4
漢字パズルコレクション 1
絵むすびコレクション

Simon Tatham's Portable Puzzle Collection



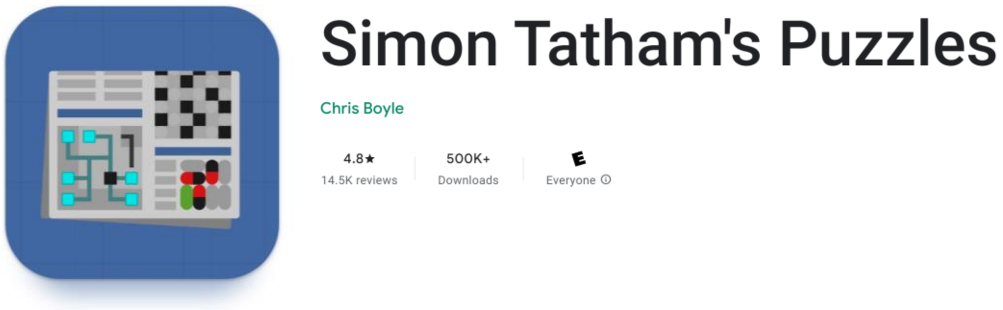
Simon Tatham's Puzzles 4+

Greg Hewgill

Designed for iPad

★★★★★ 4.8 • 171 Ratings

Free

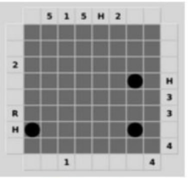


Simon Tatham's Puzzles

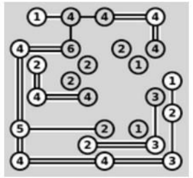
Chris Boyle

4.8★	500K+	E
14.5K reviews	Downloads	Everyone ⓘ

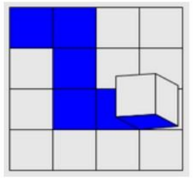
Black Box



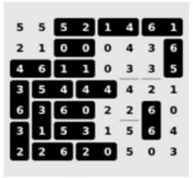
Bridges



Cube

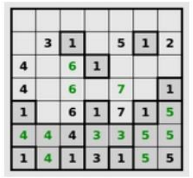


Dominosa

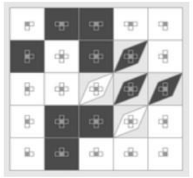


Fifteen

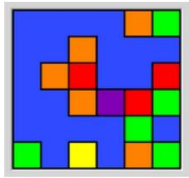
Filling



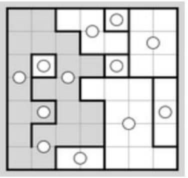
Flip



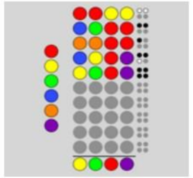
Flood



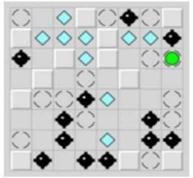
Galaxies



Guess

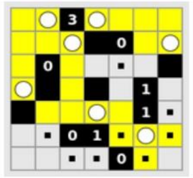


Inertia

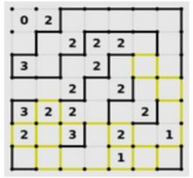


Keen

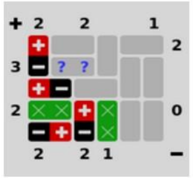
Light Up



Loopy



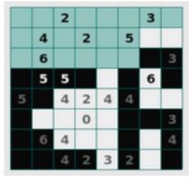
Magnets



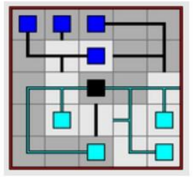
Map

Mines

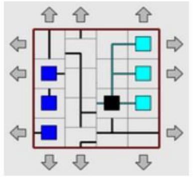
Mosaic



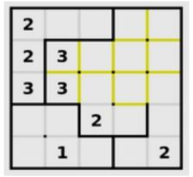
Net



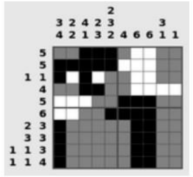
Netslide



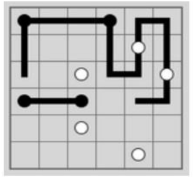
Palisade



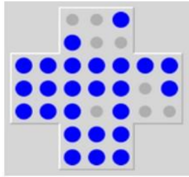
Pattern



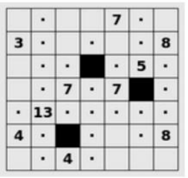
Pearl



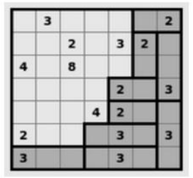
Pegs



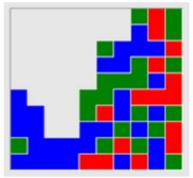
Range



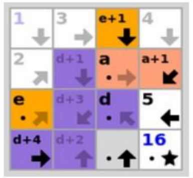
Rectangles



Same Game

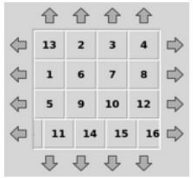


Signpost

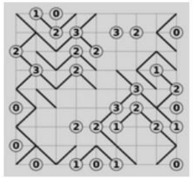


Singles

Sixteen

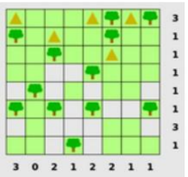


Slant

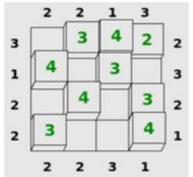


Solo

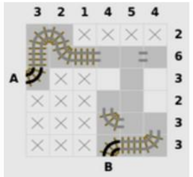
Tents



Towers

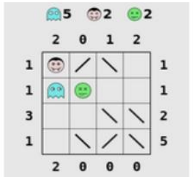


Tracks

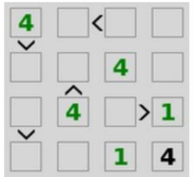


Twiddle

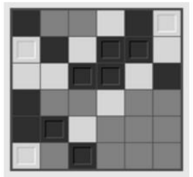
Undead



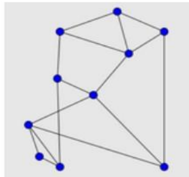
Unequal



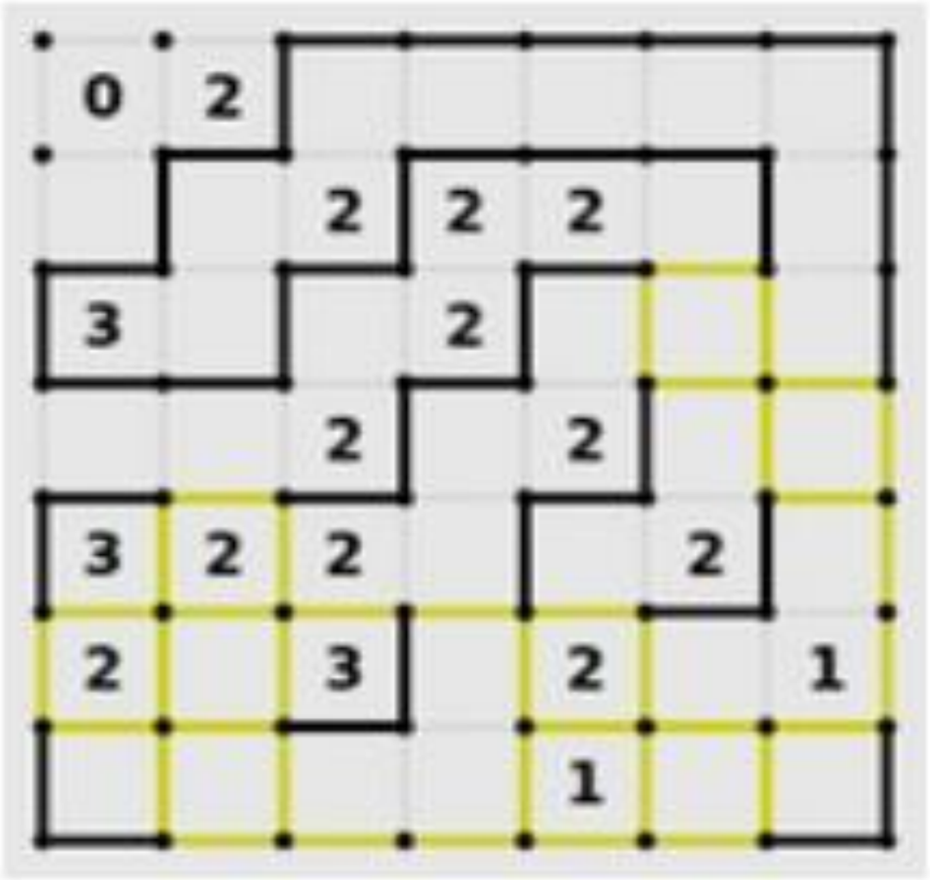
Unruly



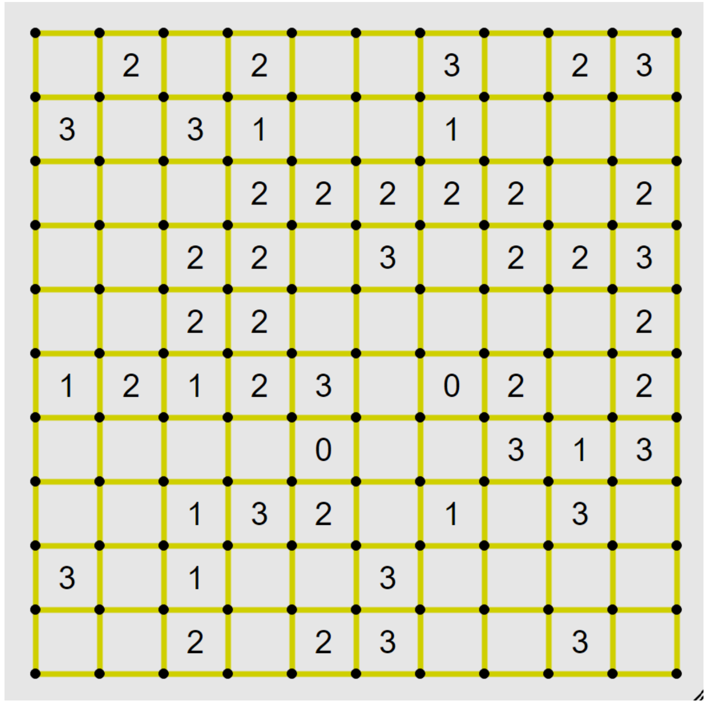
Untangle

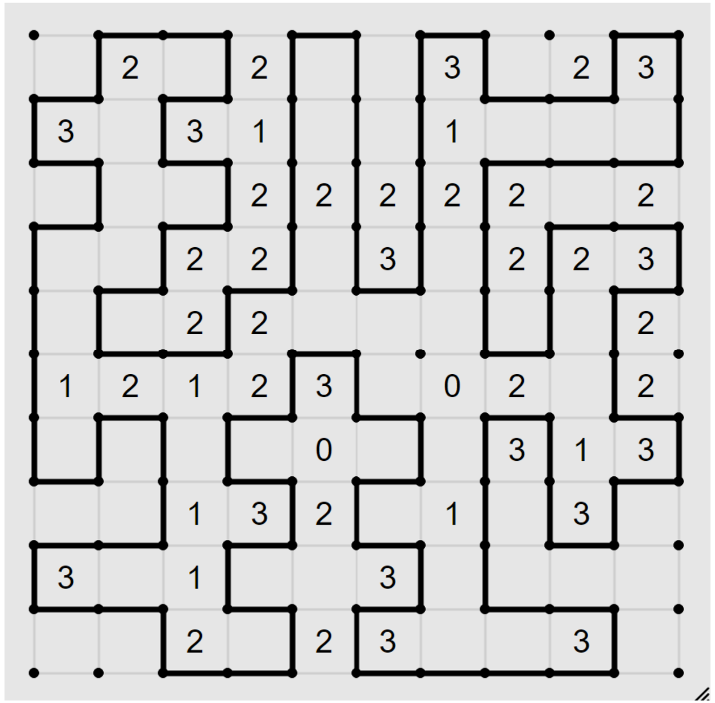


Loopy

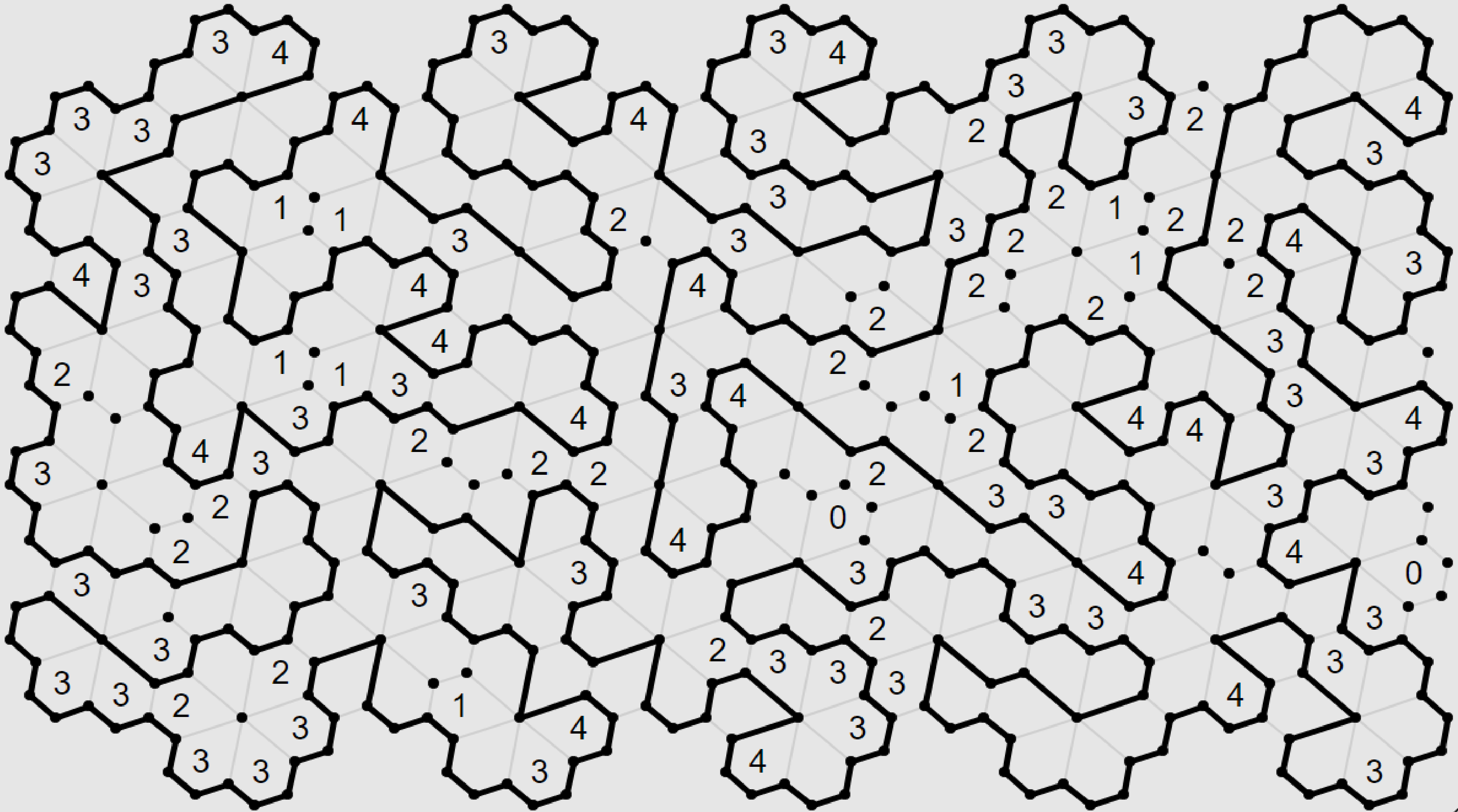


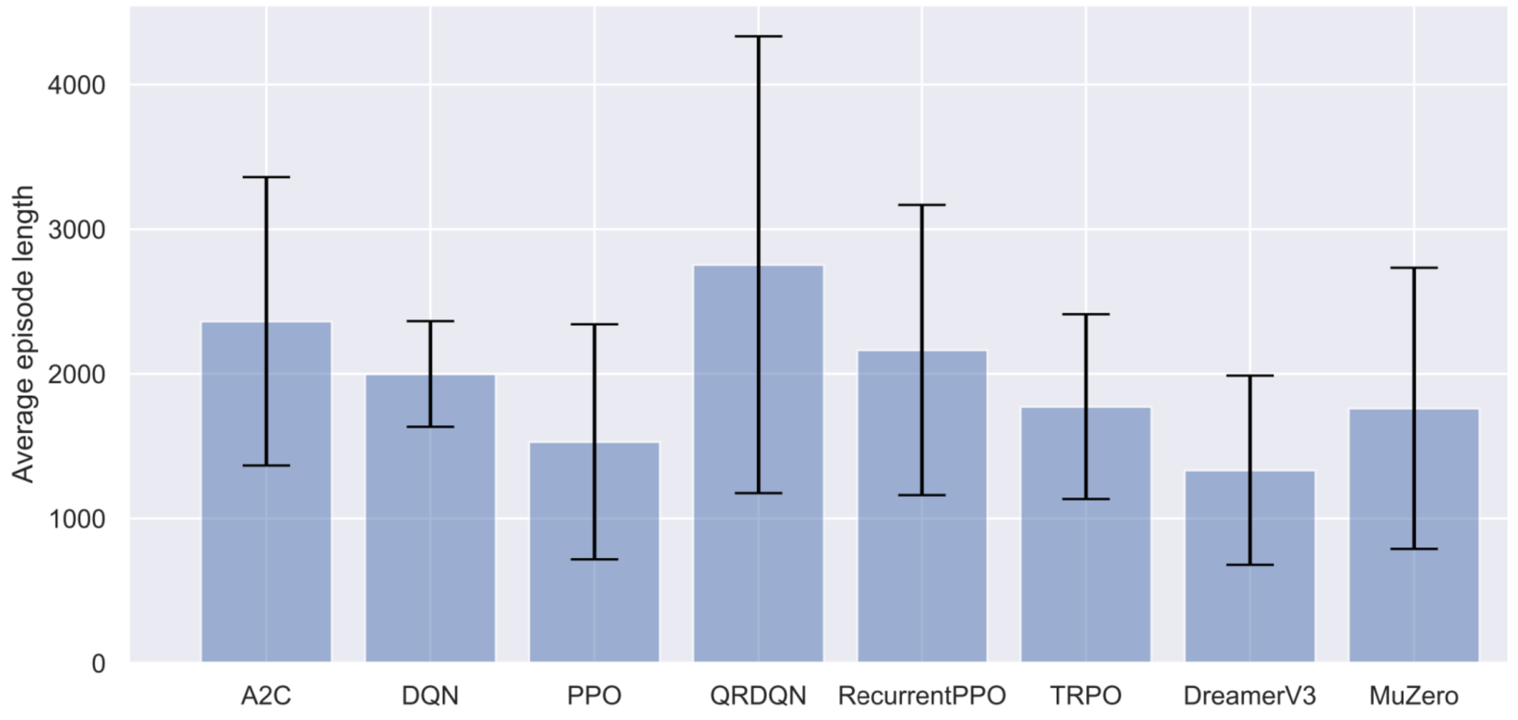
Loopy (Takegaki, Slitherlink, Ouroboros, Suriza, …)



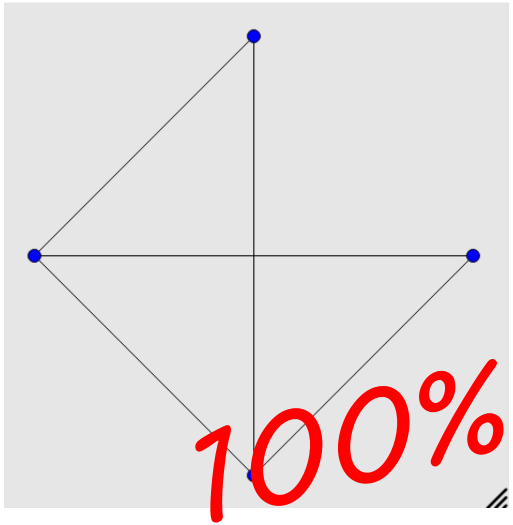


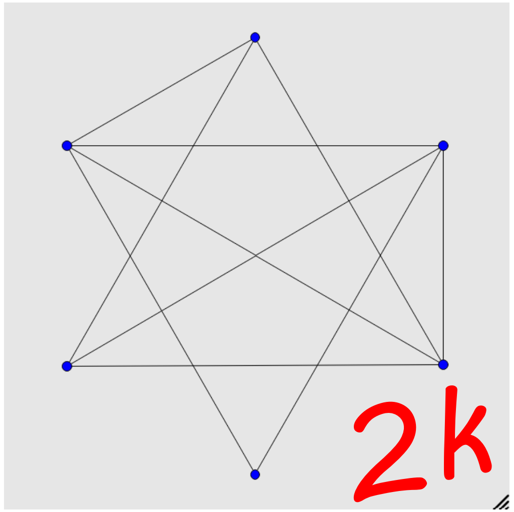
				2	2	0	2	3	1	1	2			1	2				2	1	2					0	1		2	3	1			1		0	2
		3		1	1		2				2	2	3	0			3									1				2			1	1			
3	2						2	2								2		2	3	1	2						2	0	3				2	3			
		2			1	1		1	1	1	0		2		2					1			2		1				2	1				2		1	3
	2	3		0		1	3		3		0					3			3		3	0	1		1	0	2				3	1	1	1			1
						2		2	2		1	2		2					2			1	1	2			2		3		1		3	3	3		
2	1	2					2	2	3			3			1					3		2					1		3				0				
	2	2	2		3		1		2			2													1			1	3				1		3		3
1			3			2					2	3		2	2	3	2	1	3		1		2	1	2	3	3			0				3	1		
				3				2	1	2					1		3	2		2	2						2	0		3	3						1
1	2	3	1	3					2	1		2			3				1	1		3	3	2	1		2							1	3		
				3	2		2	1		2	1	1	1	2		3		2	0	2	1		1		2	3		3				1		3			
		2	2					2		3	2						2										3		2				1	2	3		1
	3		2	2			2	2	2					2							2		1		1	2			1	2	2	0			2	1
		1	2	1	1	2	2	2	2	3									3	0	3				3	2				0			1	2		2	
		1	3	1	2	2	2	2	1		1		2		1			2		3			3	2				2	3							2	1
	2					2	2		2			2	0		1				2	2	2				2		2		2		3	3	2			2	
1	3		1	1				1				2			0	2		1					3		1		1	3	2			1			1		3
1		0	1	0	0	1	2	2	2	1	1	1		2	1			0	1	2							1	2	1		1	3		3		0	
2	3			1	1	0						3	2	2	2	1		2	2	2			3			2		3		2	2		2	1		1	
	1						3				2	1				2				2		1				1		1	2			1			3	1	

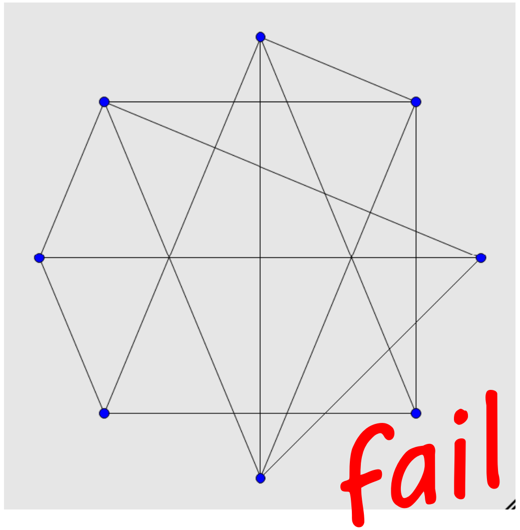



Average episode length
0
1000
2000
3000
4000
A2C
DQN
PPO
QRDQN
RecurrentPPO
TRPO
DreamerV3
MuZero

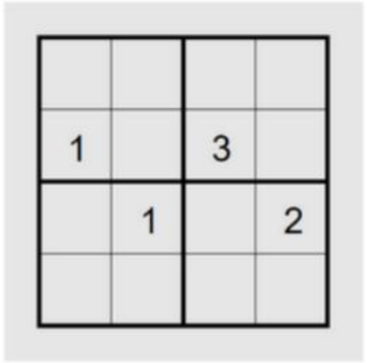
Puzzle	**Parameters**	**PPO**		**DreamerV3**	
Netslide	`2x3b1`	35.3 ± 0.7	(100.0%)	12.0 ± 0.4	(100.0%)
	`3x3b1`	4742.1 ± 2960.1	(9.2%)	3586.5 ± 676.9	(22.4%)
Same Game	`2x3c3s2`	11.5 ± 0.1	(100.0%)	7.3 ± 0.2	(100.0%)
	`5x5c3s2`	1009.3 ± 1089.4	(30.5%)	527.0 ± 162.0	(30.2%)
Untangle	`4`	34.9 ± 10.8	(100.0%)	6.3 ± 0.4	(100.0%)
	`6`	2294.7 ± 2121.2	(96.2%)	1683.3 ± 73.7	(82.0%)



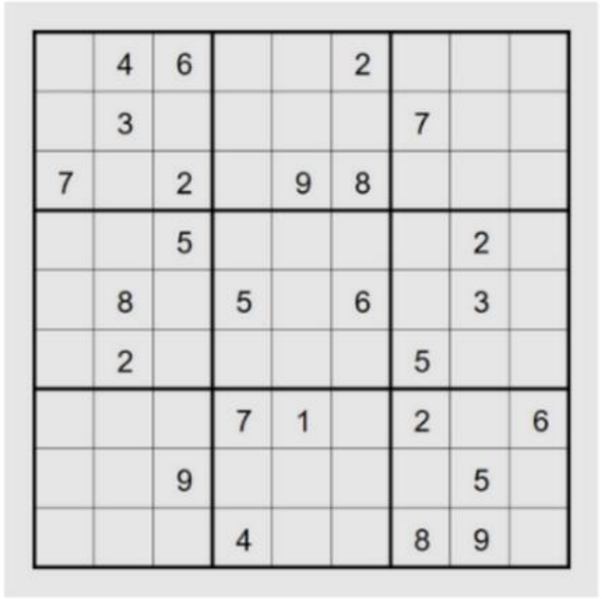




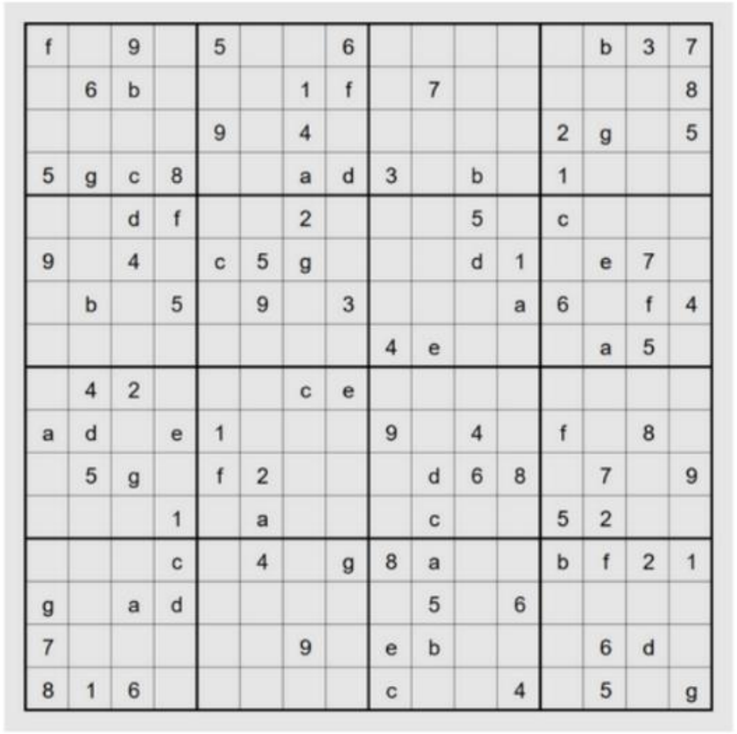
Sudoku



1		3	
	1		2

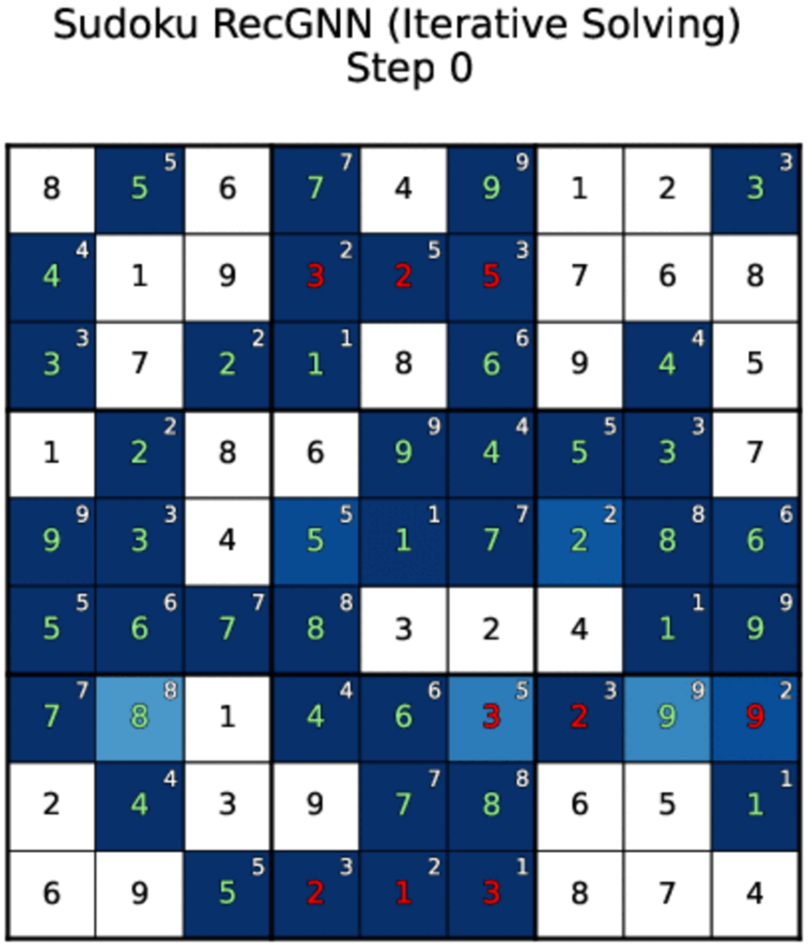


	4	6			2			
	3					7		
7		2		9	8			
		5					2	
	8		5		6		3	
	2					5		
			7	1		2		6
		9					5	
			4			8	9	

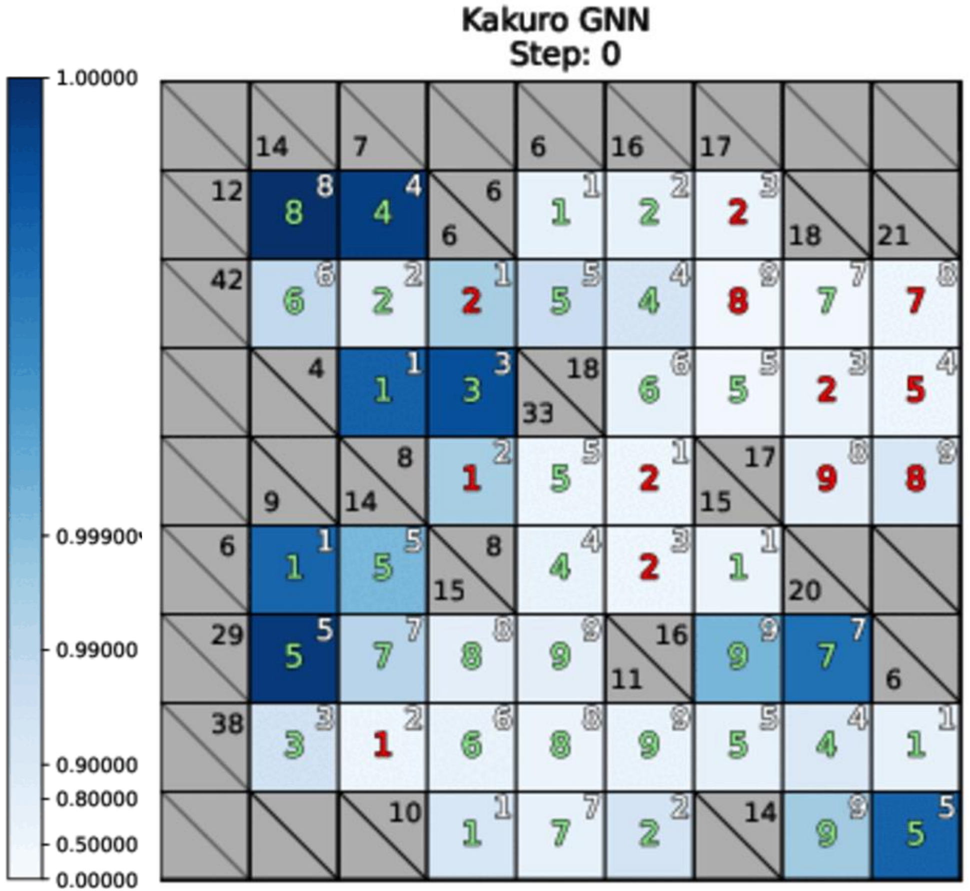


f		9		5			6						b	3	7
	6	b				1	f		7						8
				9		4						2	g		5
5	g	c	8			a	d	3		b		1			
		d	f			2				5		c			
9		4		c	5	g				d	1		e	7	
	b		5		9		3				a	6		f	4
								4	e				a	5	
	4	2				c	e								
a	d		e	1				9		4		f		8	
	5	g		f	2				d	6	8		7		9
			1		a				c			5	2		
			c		4		g	8	a			b	f	2	1
g		a	d						5		6				
7						9		e	b				6	d	
8	1	6						c			4		5		g

	4	6			2			🟥
	3					7		
7		2		9	8			
		5					2	
	8		5		6		3	
	2					5		
			7	1		2	🟩 4	6
		9					5	
			4			8	9	

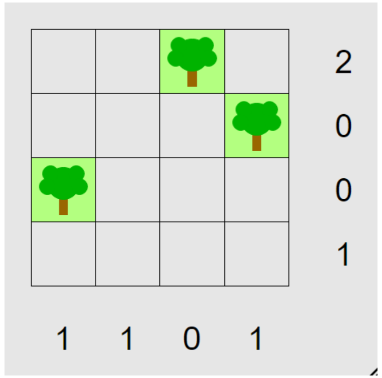


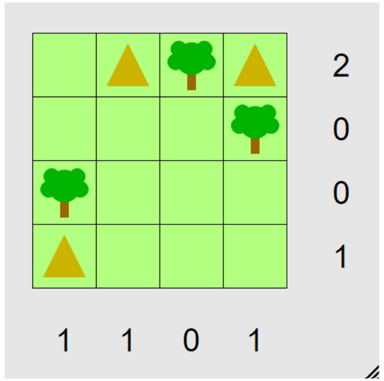
Sudoku RecGNN (Iterative Solving)
Step 0

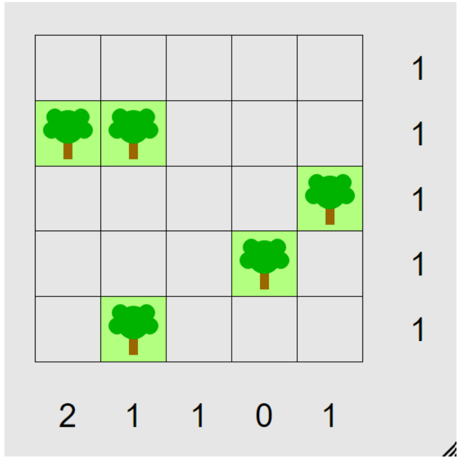


Kakuro GNN
Step: 0
1.00000
0.99900
0.99000
0.90000
0.80000
0.50000
0.00000

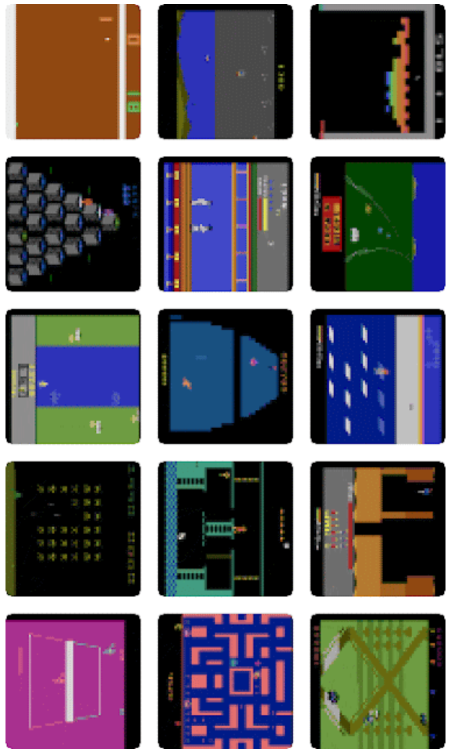
Tents

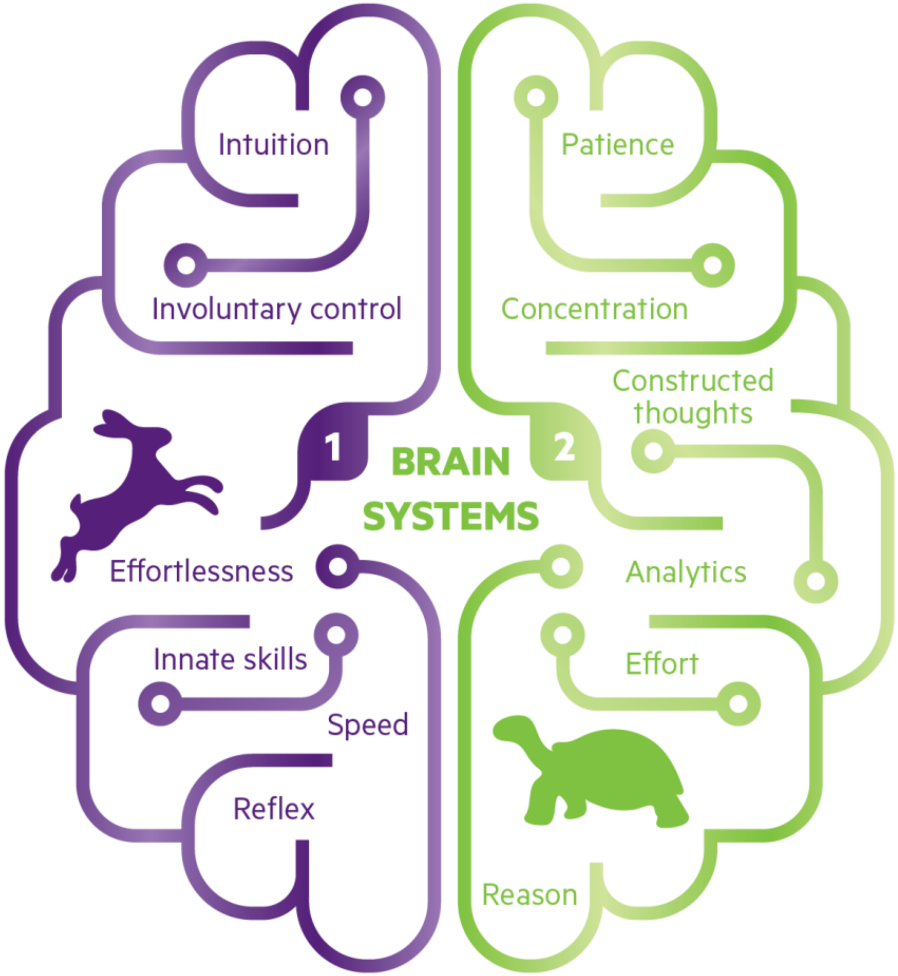




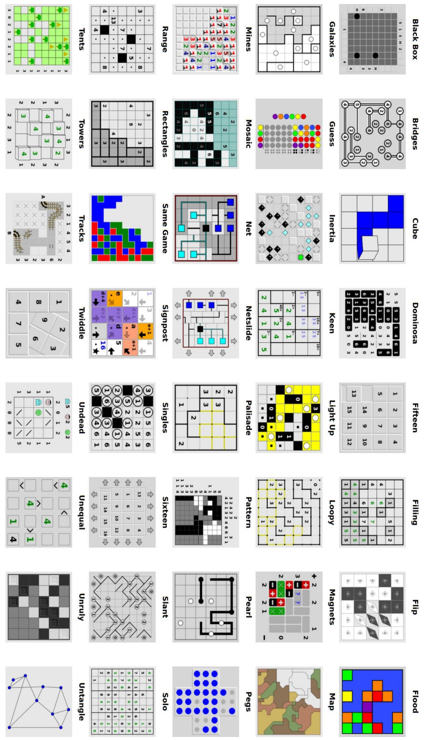


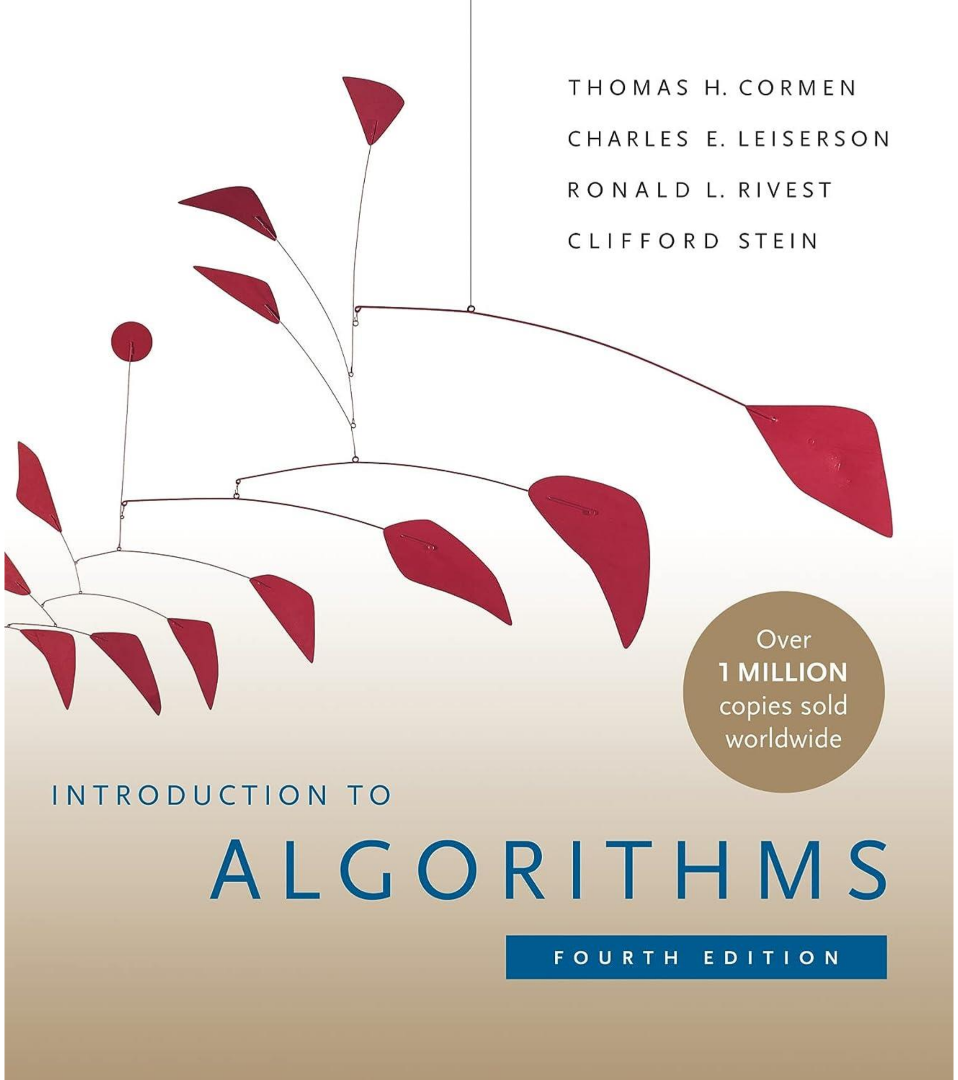
Atari Games





Puzzle Coll.





THOMAS H. CORMEN

CHARLES E. LEISERSON

RONALD L. RIVEST

CLIFFORD STEIN

Over **1 MILLION** copies sold worldwide

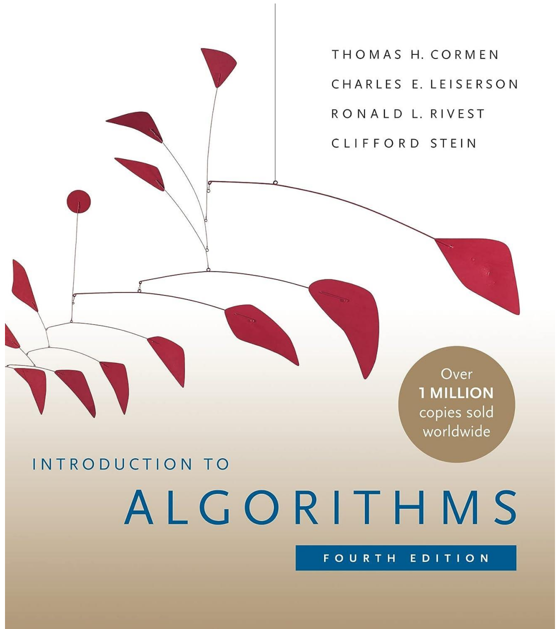
INTRODUCTION TO

ALGORITHMS

FOURTH EDITION

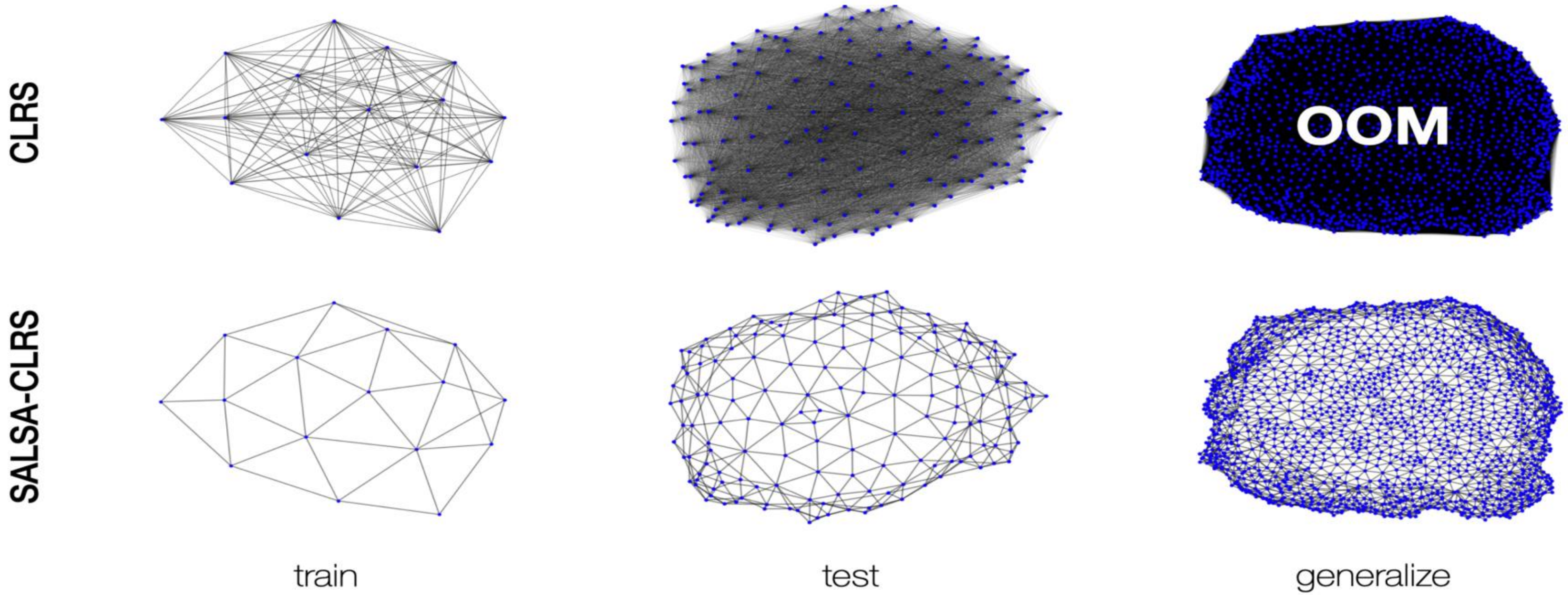
The CLRS Algorithmic Reasoning Benchmark

Petar Veličković [1] Adrià Puigdomènech Badia [1] David Budden [1]
Razvan Pascanu [1] Andrea Banino [1] Misha Dashevskiy [1] Raia Hadsell [1] Charles Blundell [1]



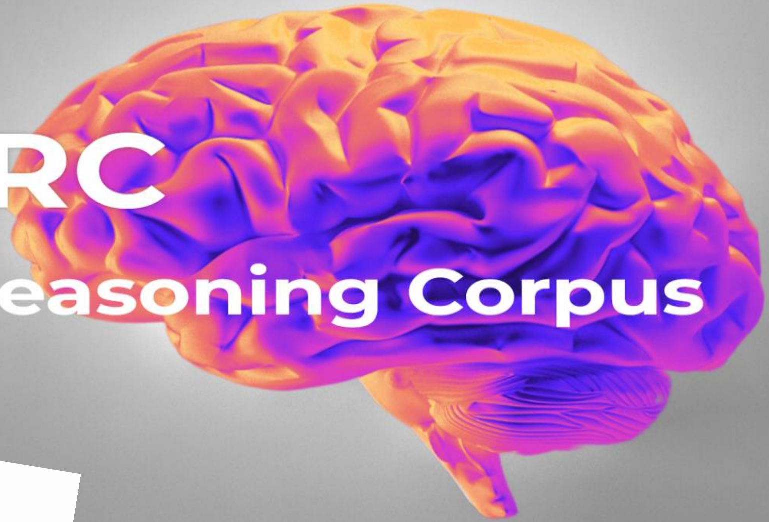
SALSA-CLRS: A Sparse and Scalable Benchmark for Algorithmic Reasoning

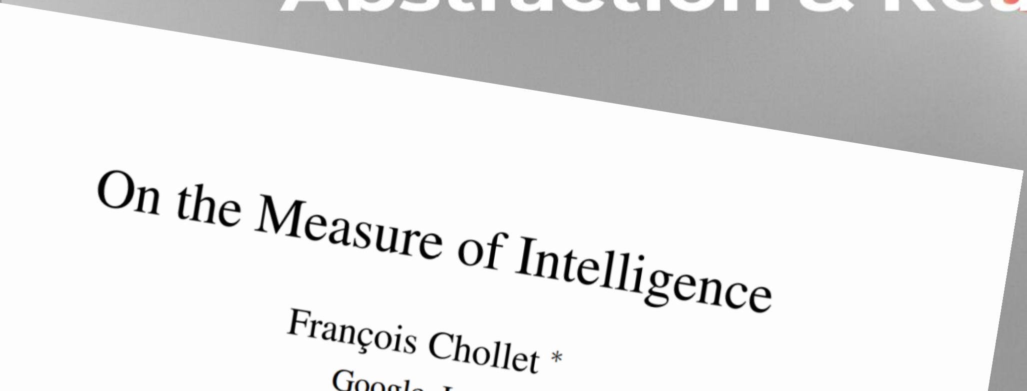
Julian Minder   Florian Grötschla*   Joël Mathys*   Roger Wattenhofer



ARC

Abstraction & Reasoning Corpus

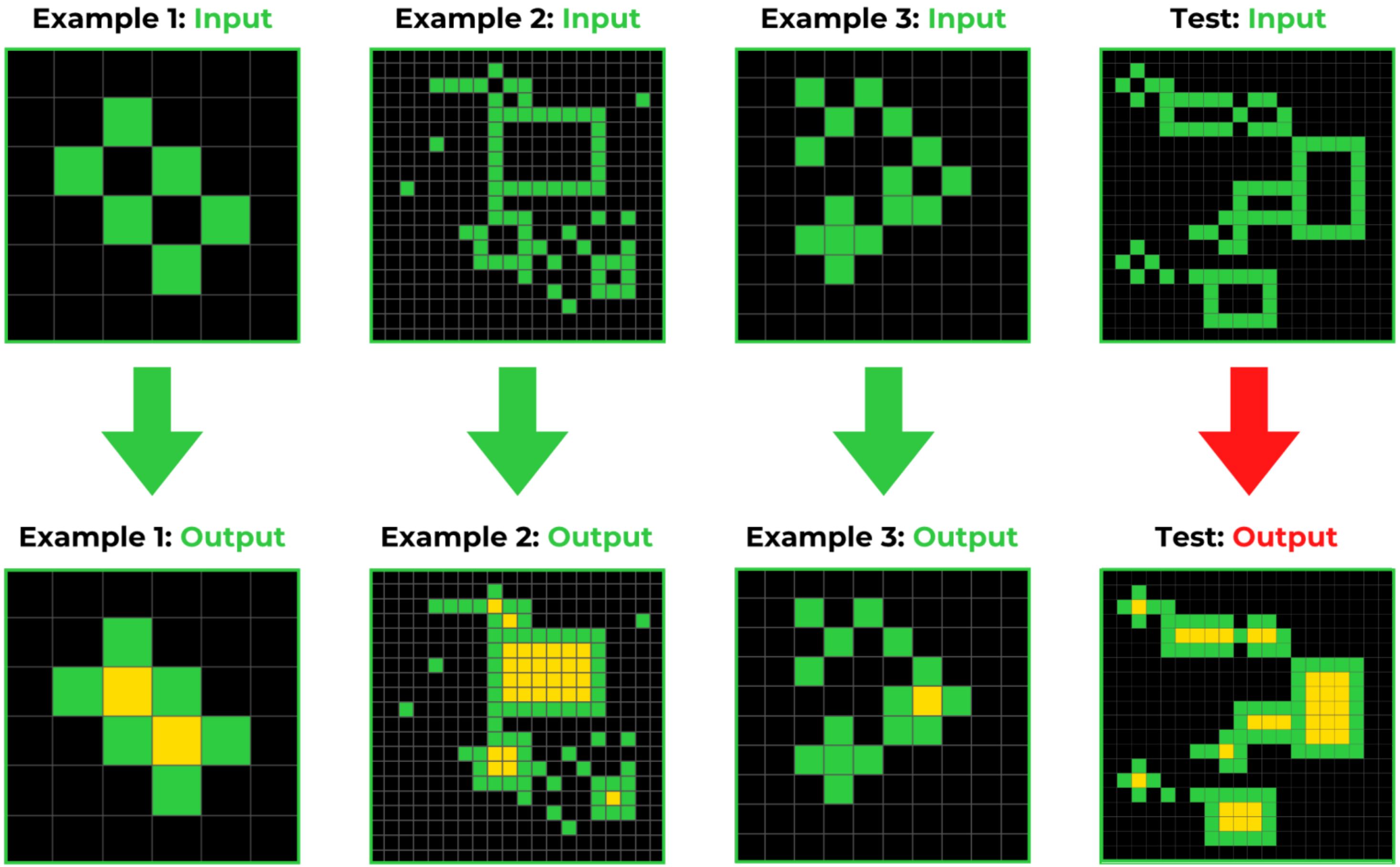


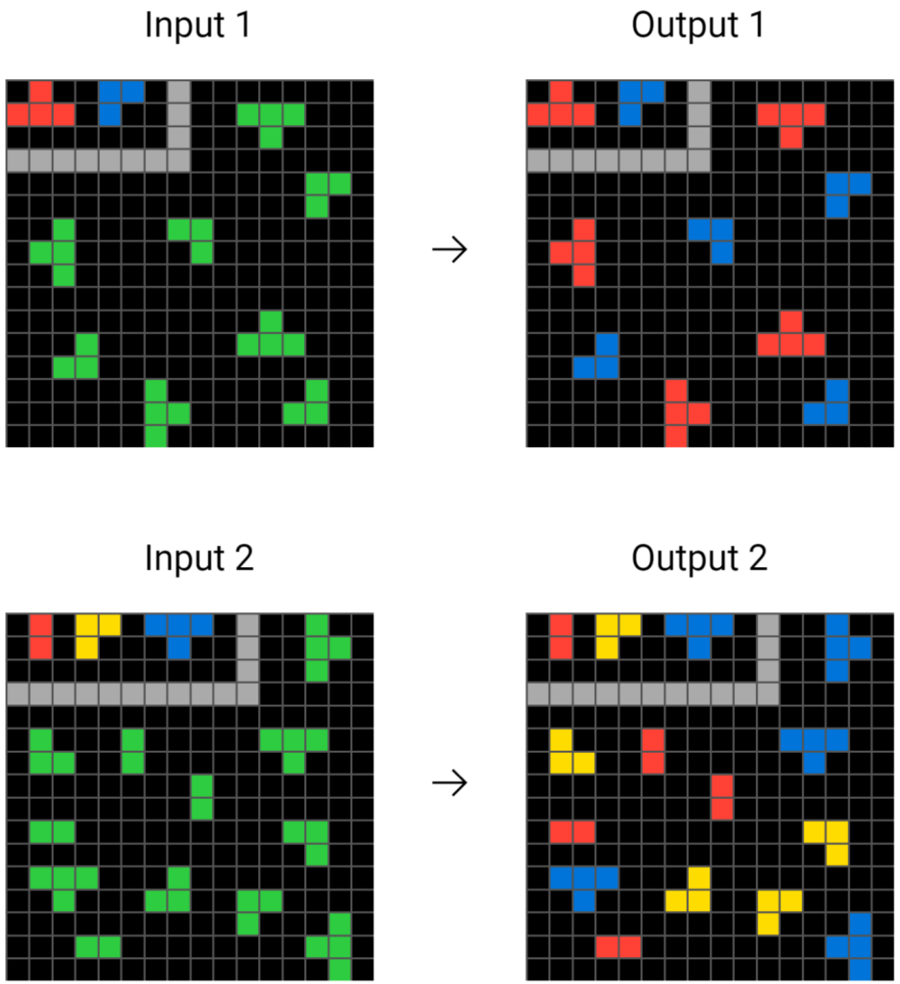


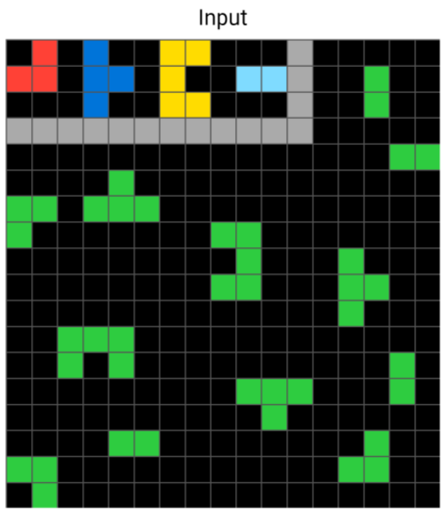
On the Measure of Intelligence

François Chollet *

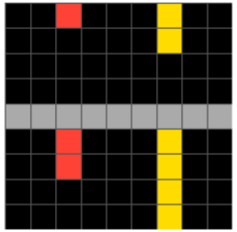
Google


Example 1: Input
Example 2: Input
Example 3: Input
Test: Input
Example 1: Output
Example 2: Output
Example 3: Output
Test: Output


Input 1
Output 1
Input 2
Output 2

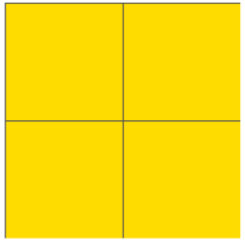

Input

Input 1

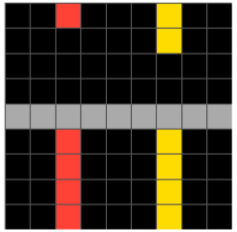


→

Output 1

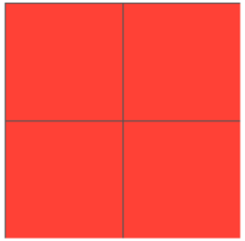


Input 2

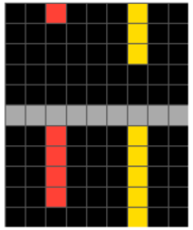


→

Output 2

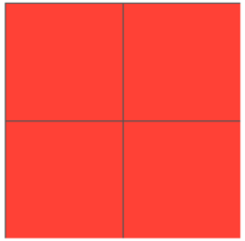


Input 3

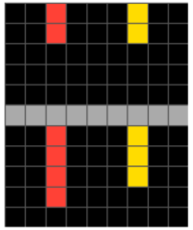


→

Output 3

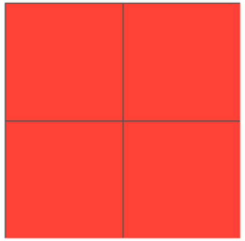


Input 4

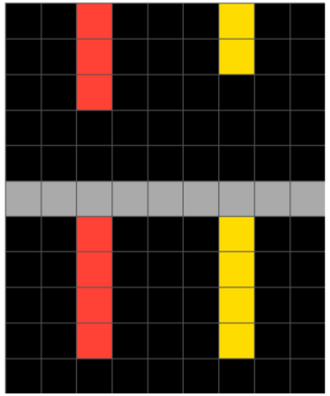


→

Output 4

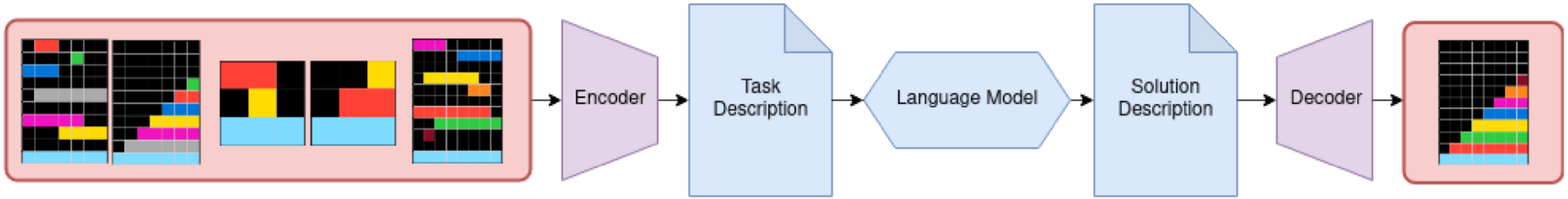


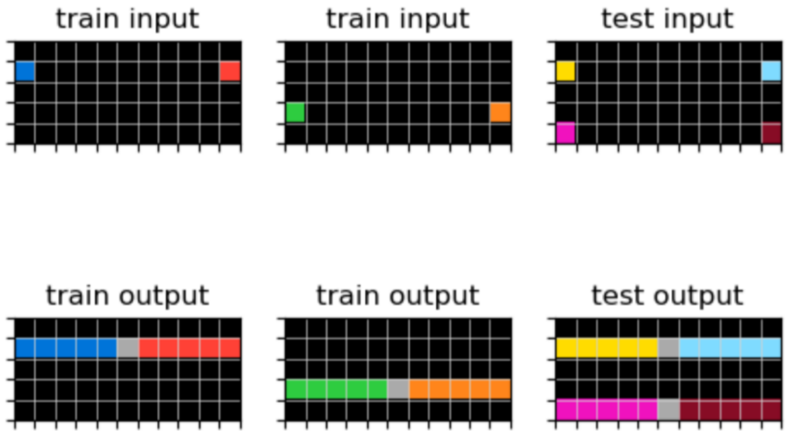
Input

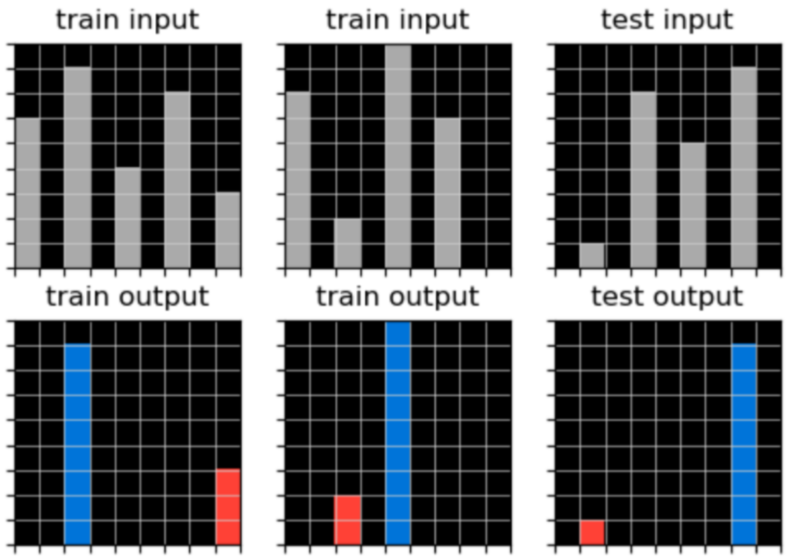


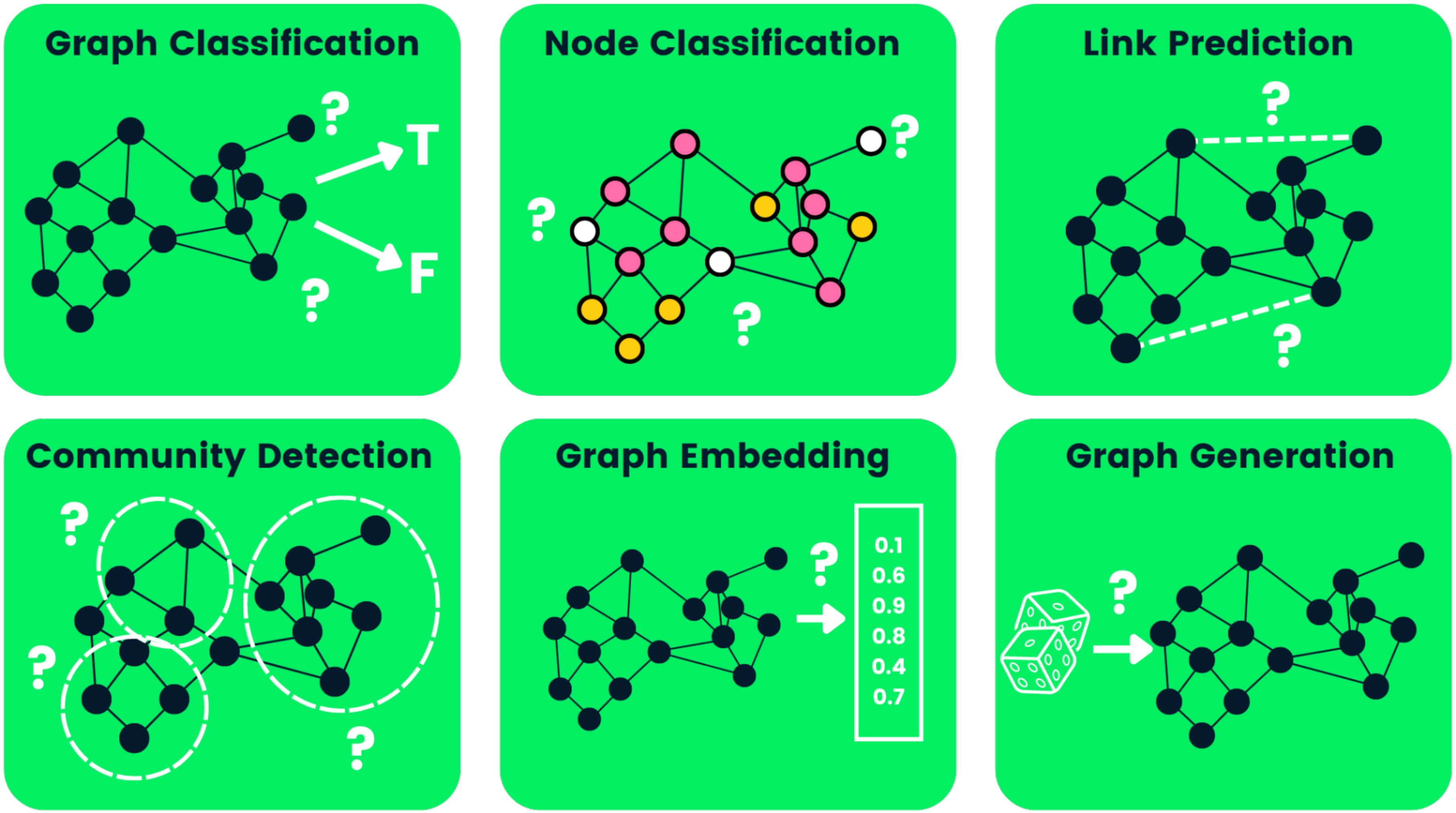
Abstract Visual Reasoning Enabled by Language

Giacomo Camposampiero* Loïc Houmard* Benjamin Estermann Joël Mathys
Roger Wattenhofer
ETH Zürich, Switzerland
{gcamposampie, lhoumard, estermann, jmathys, wattenhofer}@ethz.ch

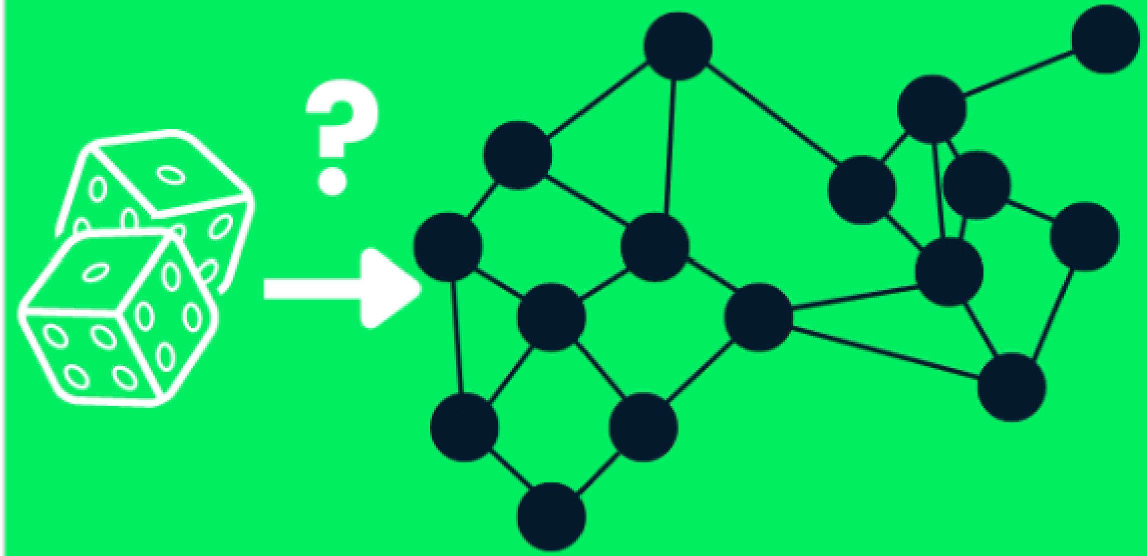



train input
train input
test input
train output
train output
test output


train input
train input
test input
train output
train output
test output

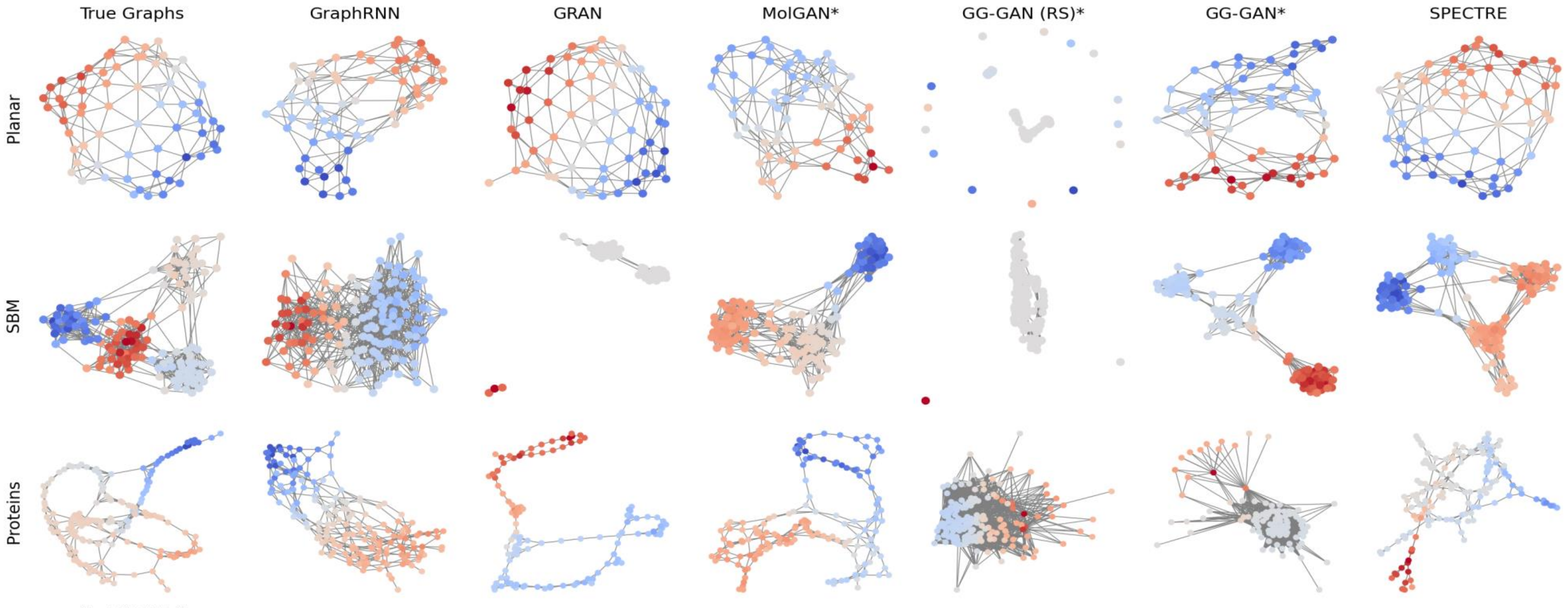

Graph Classification
?
T
F
?
Node Classification
?
?
?
Link Prediction
?
?
Community Detection
?
?
?
Graph Embedding
?
0.1
0.6
0.9
0.8
0.4
0.7
Graph Generation
?

Graph Generation



SPECTRE : Spectral Conditioning Helps to Overcome the Expressivity Limits of One-shot Graph Generators

Karolis Martinkus [1] Andreas Loukas [*2] Nathanaël Perraudin [*3] Roger Wattenhofer [1]

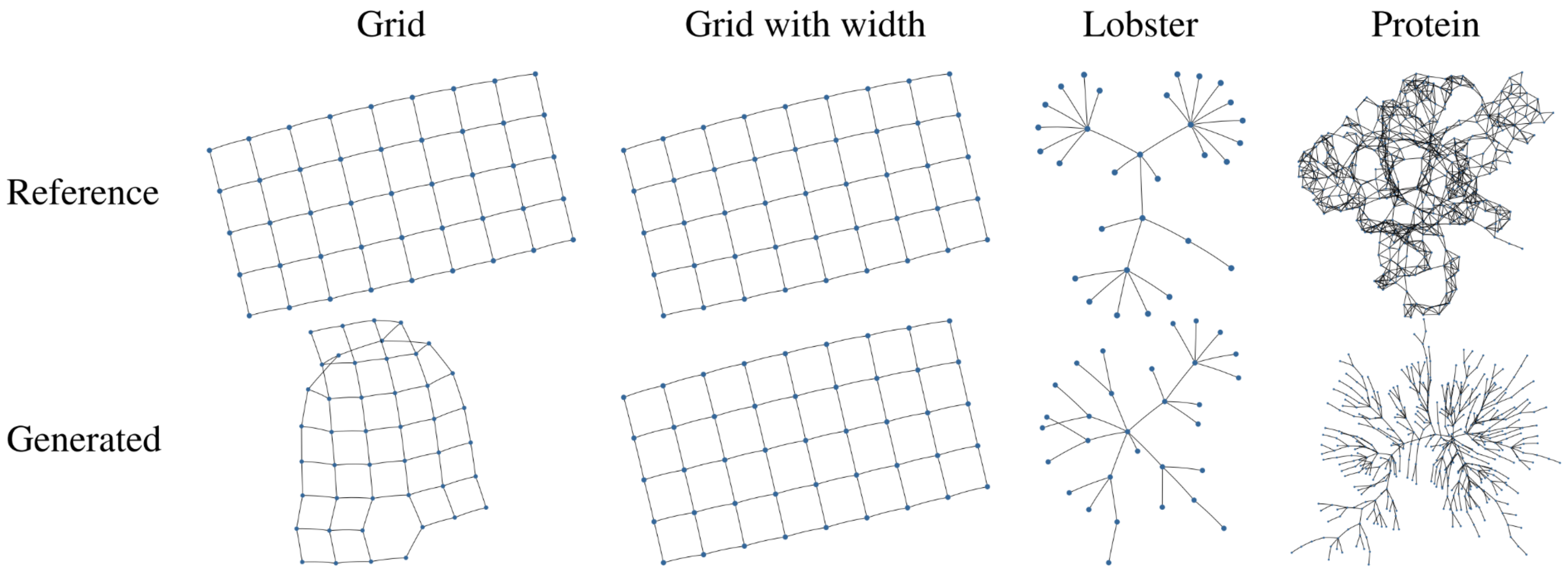


Discovering Graph Generation Algorithms

Mihai Babiac, Karolis Martinkus & Roger Wattenhofer
ETH Zurich
{mbabiac,martinkus,wattenhofer}@ethz.ch

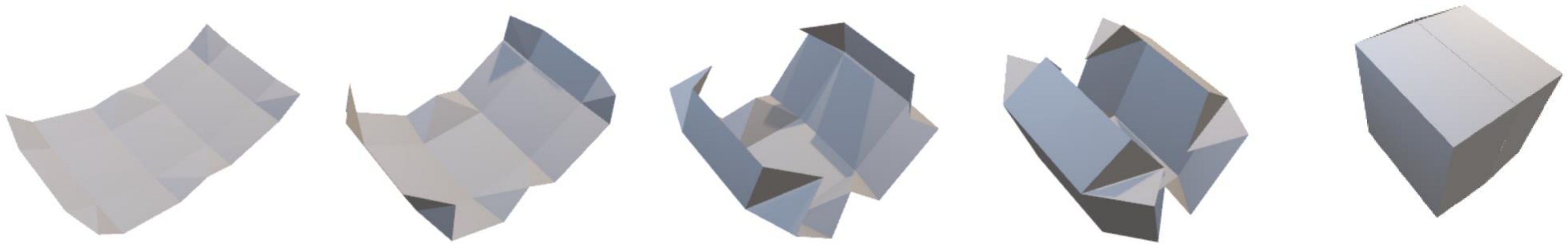
```python
def outer_loop():
    for i in range(N):
        inner_loop()

def inner_loop():
    for j in range(i):
        float00 = random(0, 1)
        bool00 = float00 < 0.4
        if bool00:
            add_edge(i, j)

outer_loop()
```

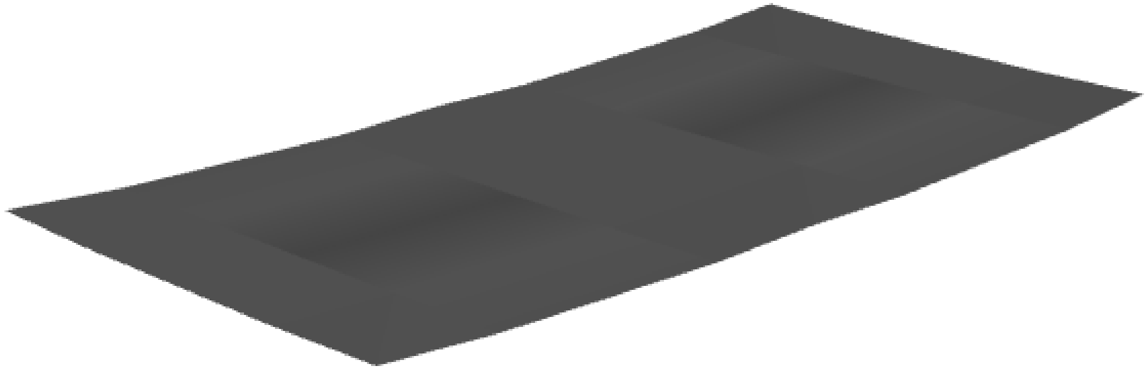
```python
def outer_loop():
    for i in range(N):
        int00 = i + n
        add_edge(i, int00)

        int01 = i % n
        bool00 = int01 == 0
        if not bool00:
            int01 = i + 1
            add_edge(i, int01)

outer_loop()
```

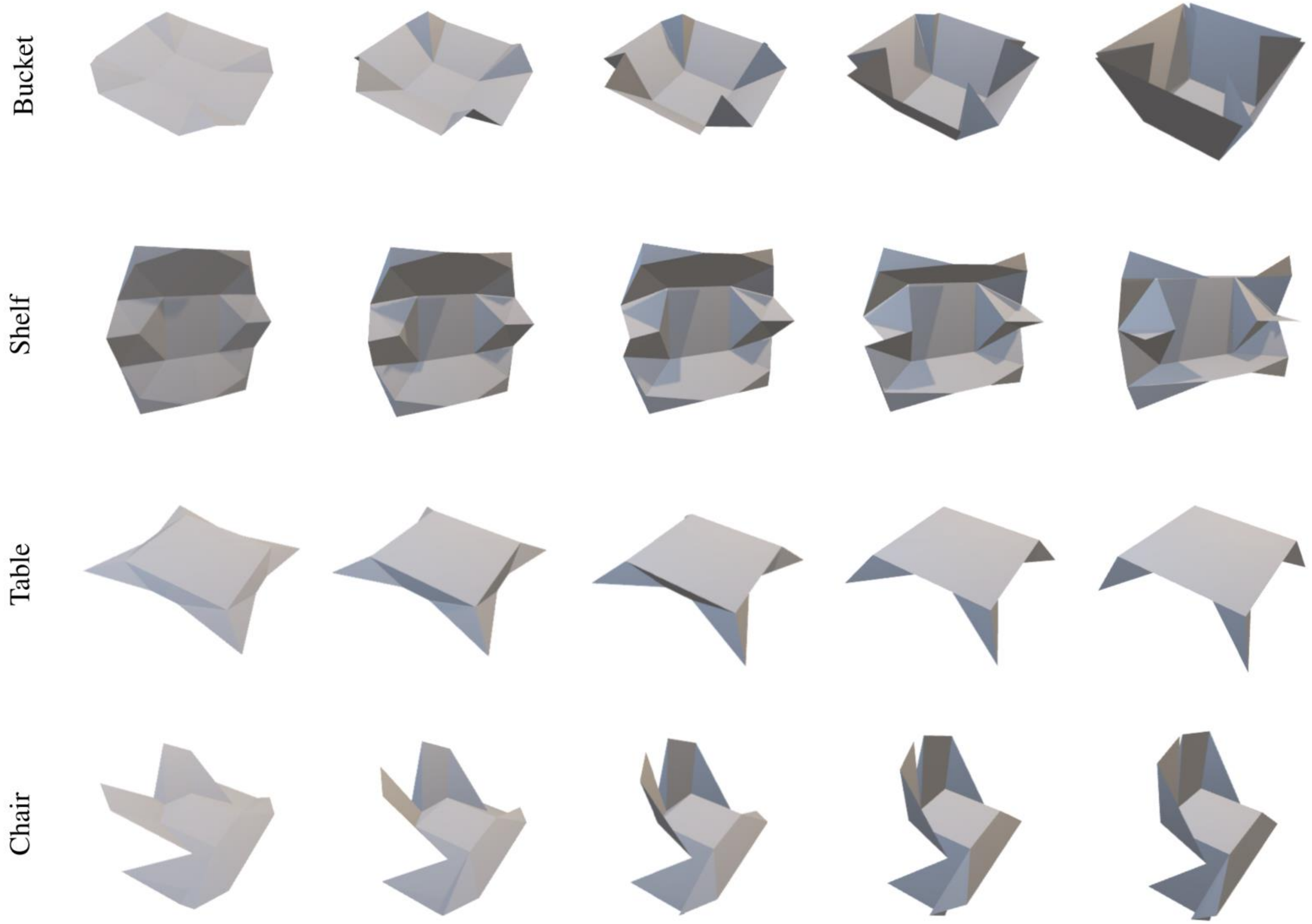

Grid
Grid with width
Lobster
Protein
Reference
Generated

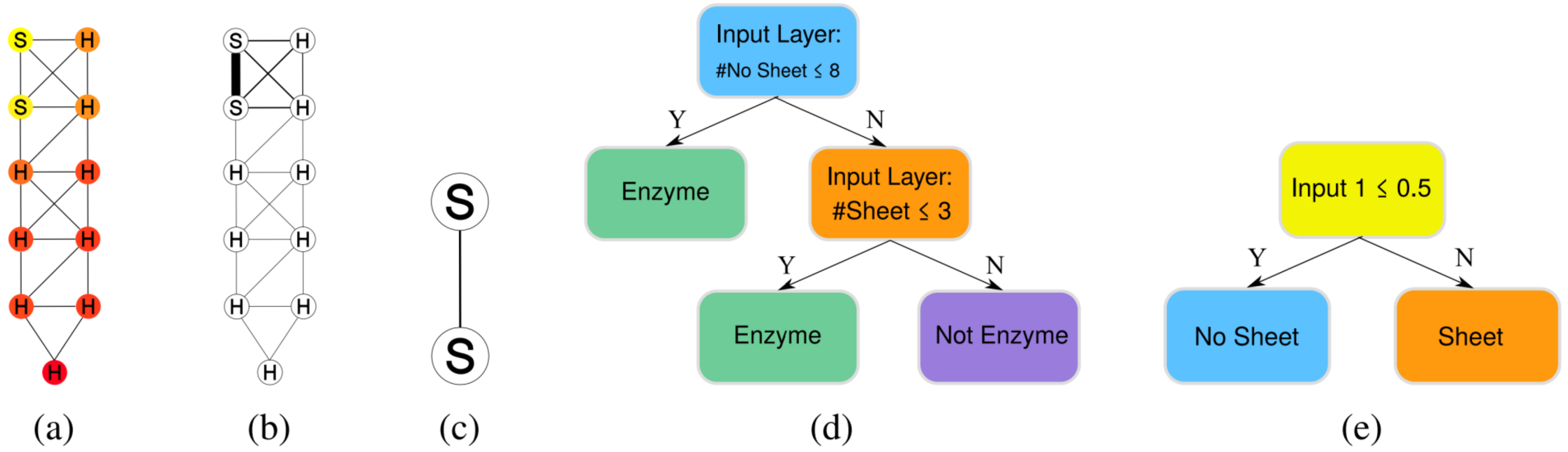
Automating Rigid Origami Design

Jeremia Geiger, Karolis Martinkus, Oliver Richter, Roger Wattenhofer



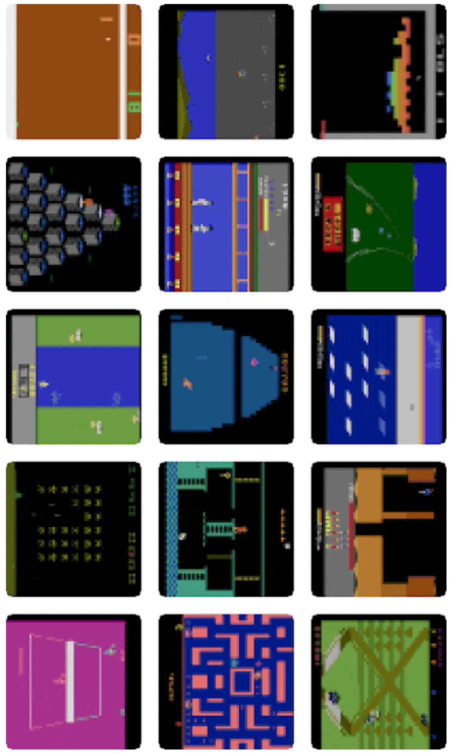


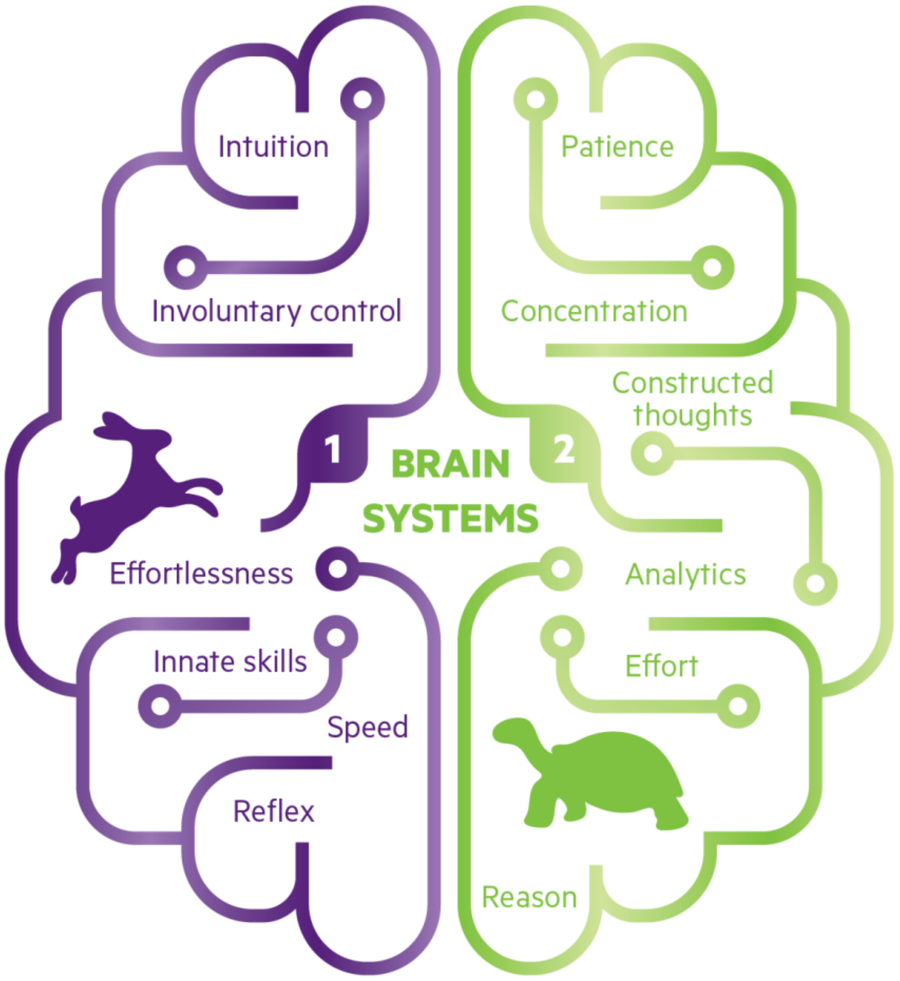

Bucket
Shelf
Table
Chair


S
H
(a)
(b)
(c)
Input Layer: #No Sheet ≤ 8
Y
N
Enzyme
Input Layer: #Sheet ≤ 3
Y
N
Enzyme
Not Enzyme
(d)
Input 1 ≤ 0.5
Y
N
No Sheet
Sheet
(e)

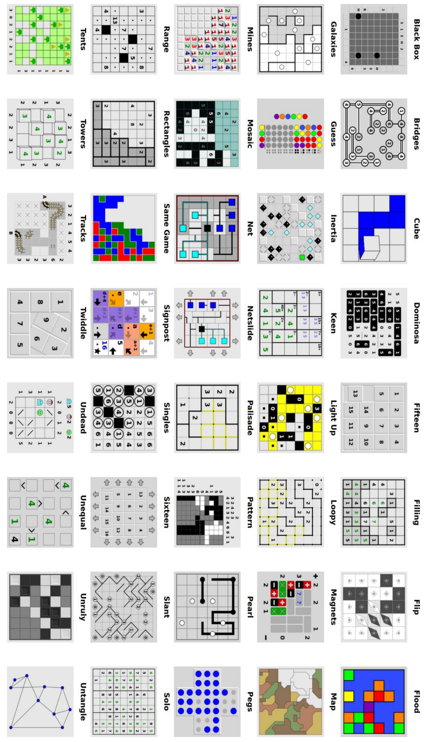
The Bigger Picture

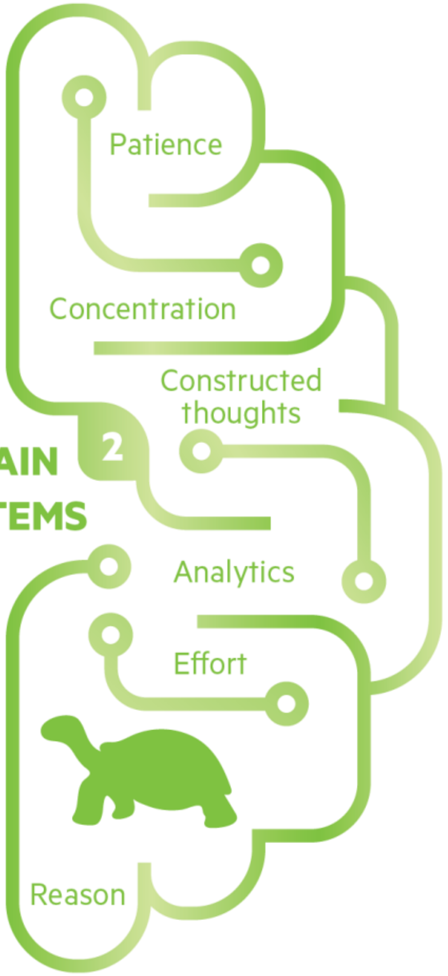
Atari Games





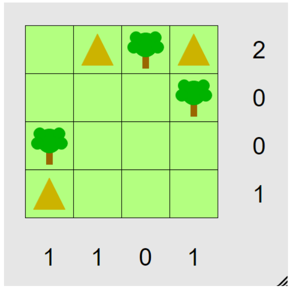
Puzzle Coll.

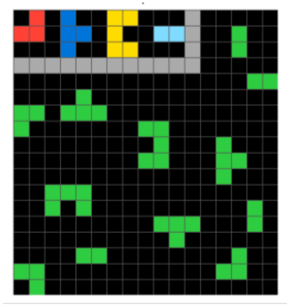


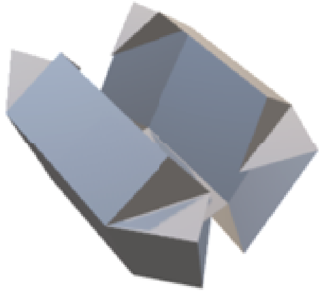


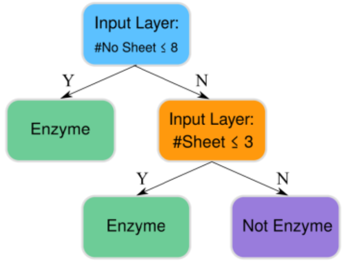
Problem = Graph

Solution = Dist. Learning









Thank You!

Any questions or comments?

Thanks to co-authors: Peter Belcák, Benjamin Estermann, Lukas Faber, Florian Grötschla, Ard Kastrati, Luca Lanzendörfer, Karolis Martinkus, Joël Mathys, etc.

Roger Wattenhofer