6th Strasbourg Summer School in Chemoinformatics (CS3-2018)

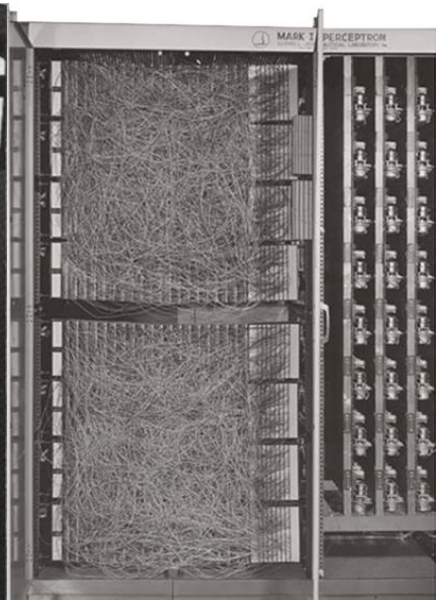
Deep Learning: What Makes Neural Networks Great Again?

Igor I. Baskin

M. V. Lomonosov Moscow State University

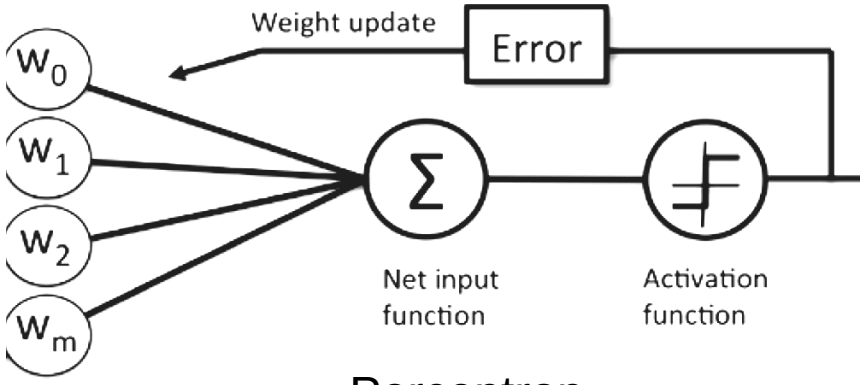
The First Neural Network (Perceptron) and the First Neurocomputer (Mark I), 1957

Frank Rosenblatt
(1928-1971)



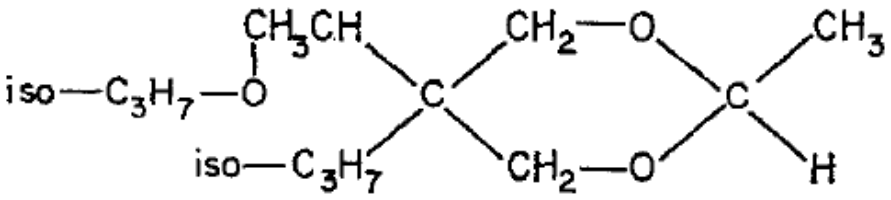
Mark I Perceptron

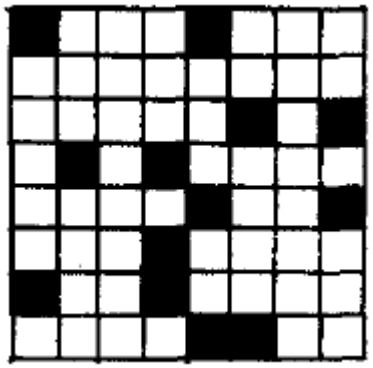
- Designed for image recognition
- “Knowledge” is encoded as weights in potentiometers (variable resistors)
- Weights are updated during learning performed by electric motors



Perceptron

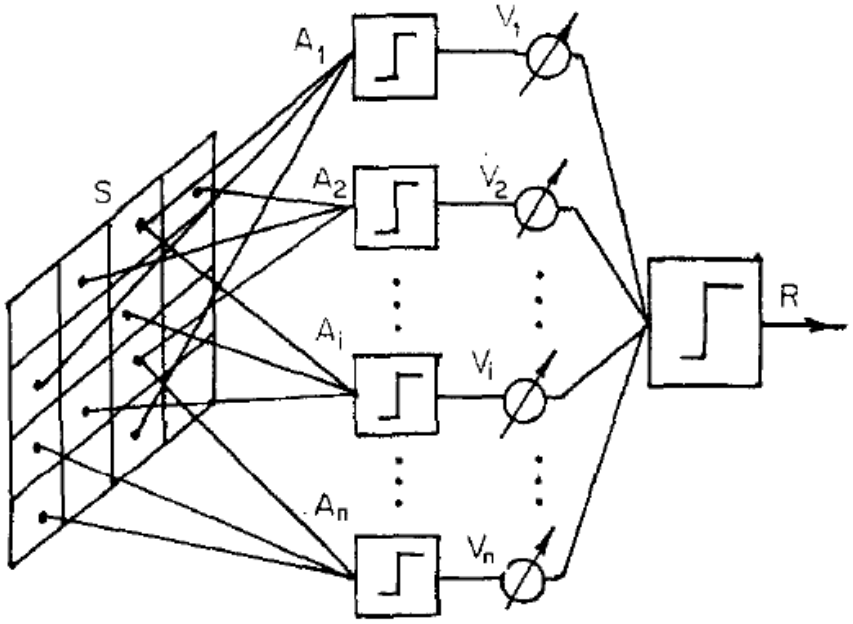
The First Use of Neural Networks (Multilayer Perceptron) in Chemoinformatics (1971)





Projection of structural formula onto retina (2D fingerprint)

Prediction of anticonvulsant activity of 1,3-dioxanes (antagonism to corasol)



A three-layer perceptron

S.A. Hiller et al. *Doklady Akademii Nauk*, **1971**, vol. 199, pp. 851-853 (Russ.);
S.A. Hiller et al. *Computers and Biomedical Research*, **1973**, vol. 6, pp. 411-421.

The First Period of Disappointment with Neural Networks (1972-1986)

Marvin Lee Minsky
(1927-2016)

Seymour Papert
(1928-2016)

XOR truth table

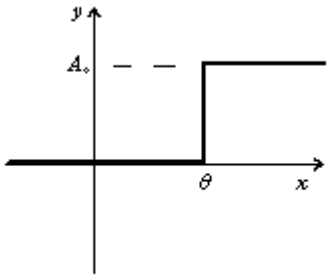
Input		Output
A	B	
0	0	0
0	1	1
1	0	1
1	1	0

- 0, false
- 1, true

Marvin Minsky and Seymour Papert, 1972 (2nd edition with corrections, first edition 1969) Perceptrons: An Introduction to Computational Geometry, The MIT Press, Cambridge MA

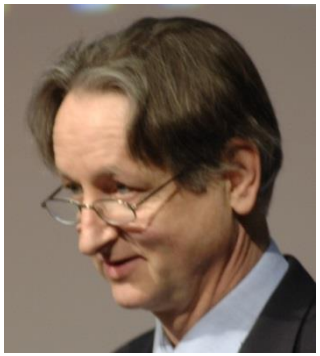
(Re)Discovery of Backpropagation Algorithm (1986)

- By replacing the threshold activation function with the sigmoid function and applying the backpropagation algorithm, it is possible to train multilayer neural networks with one or several (few) hidden layers
- Such networks can approximate any dependence between variables, including the XOR function

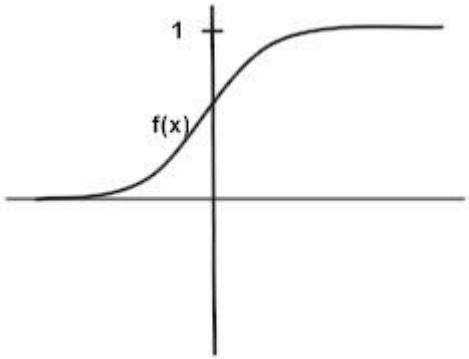


David Rummelhard

James McClelland

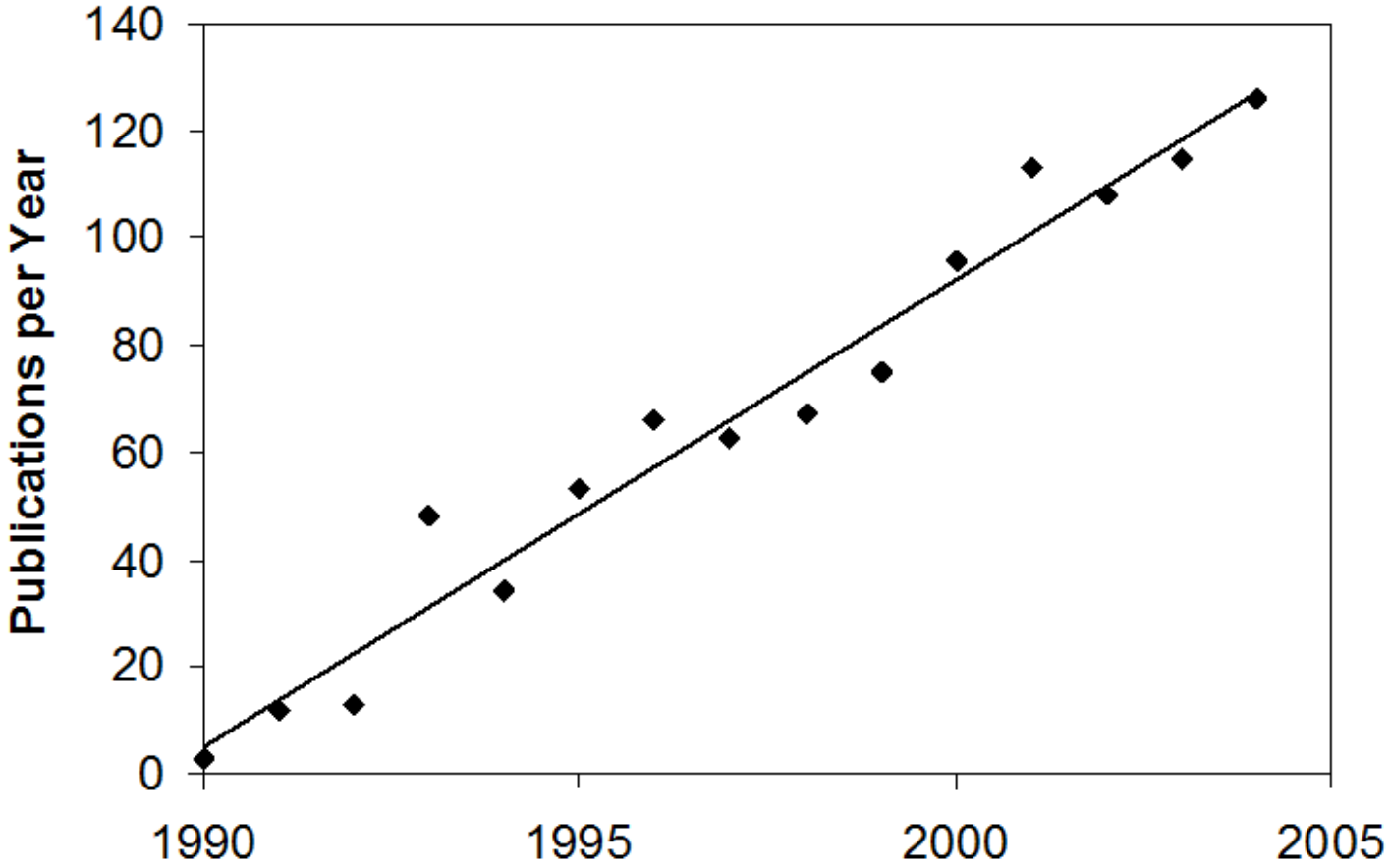


Geoffrey Hinton



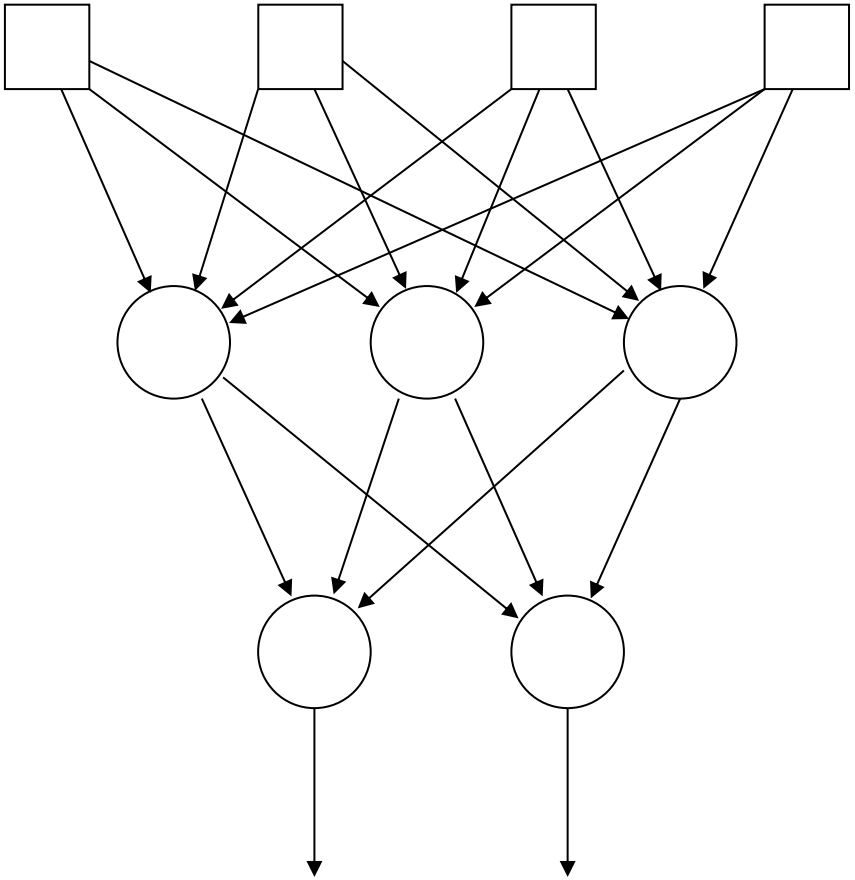
- D.E. Rumelhart, J.L. McClelland, *Parallel Distributed Processing*. Vol. 1,2. **1986**, Cambridge, MA: MIT Press
- D.E. Rumelhart, G.E. Hinton, R.J. Williams, *Nature*, **1986**, v. 33. pp. 533-536

The “Linear Growth Law” for the Number of Papers Published per Year on the Use of Neural Networks in Chemoinformatics [2]



[1] N.Halberstam, et al. *Russ. Chem. Rev*. 72 (**2003**) 629-649
[2] I.I.Baskin, et al. *Methods Mol. Biol*. 458 (**2008**) 137-158
[3] I.I.Baskin, et al.. *Expert Opin. Drug Discov.* 11 (**2016**) 785-795

“Shallow” Multilayer Feed-Forward Neural Network



Input layer
(descriptors)

Hidden layer
(non-linear latent variable)

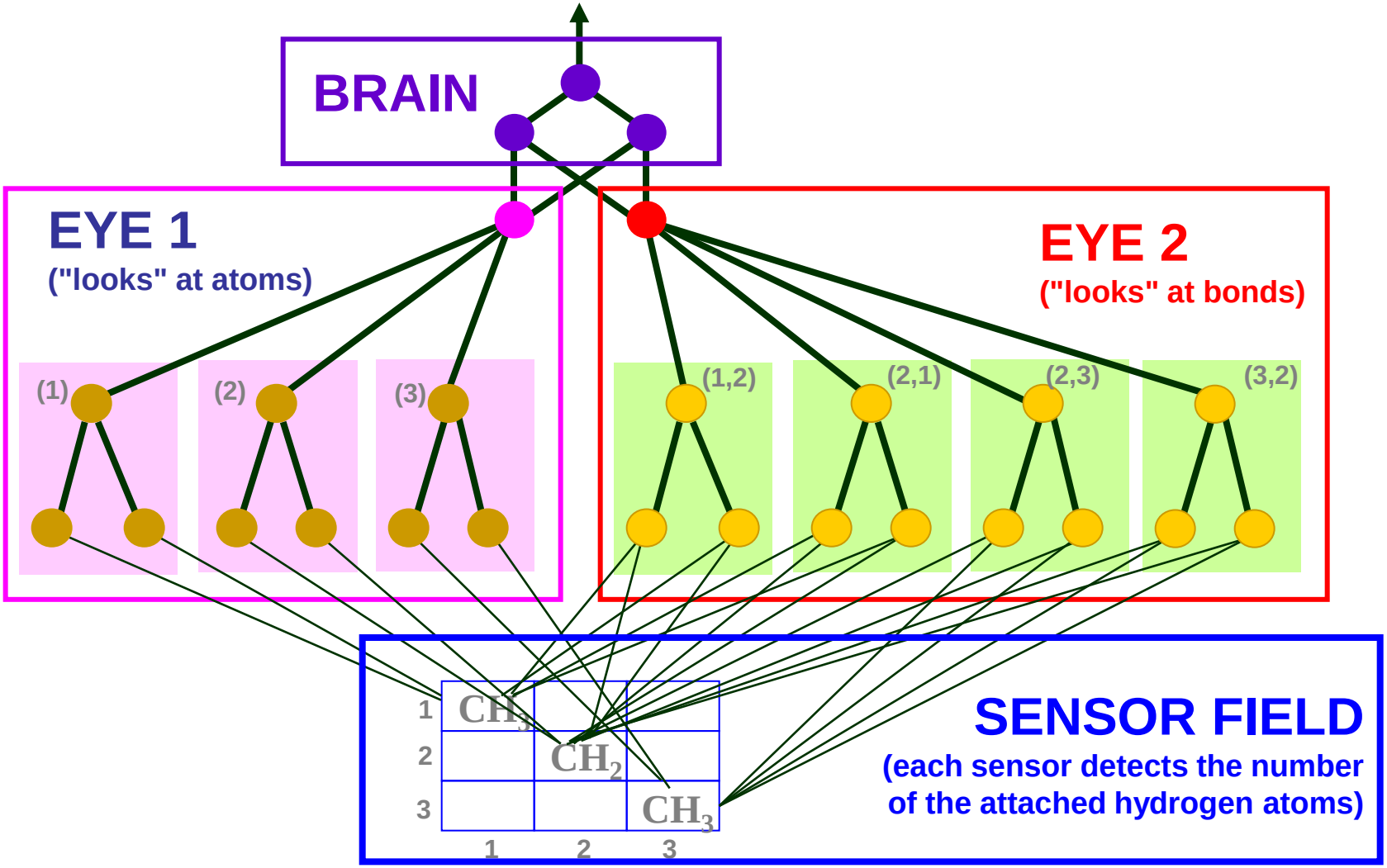
Output layer
(properties being predicted)

Any mapping from molecular graphs to properties can be approximated with the combination of such neural networks with fragment descriptors

(Baskin et al., *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 527-533)

Neural Device for Searching Direct Correlations between Structures and Properties of Chemical Compounds

Neural device in application to the propane molecule :



Baskin I.I. et al, *J. Chem. Inf. Comput. Sci*, **1997**, *37*, 715-721

The Second Period of Disappointment with Neural Networks (1997-2010)

Statistical learning theory
Support vector machines (SVM)

Vladimir Vapnik
(b. 1936)

Decision trees
Ensemble modeling
Random forest (RF)

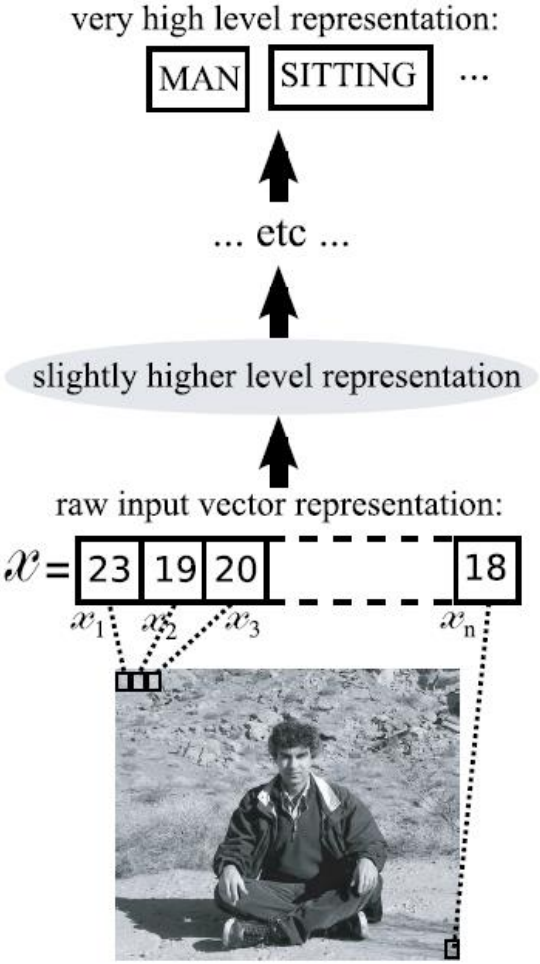
Leo Breiman
(1928-2005)

- Higher computational efficiency
- Comparable or higher predictive performance
- Strong mathematical foundation
- Models are better and easier interpretable

Deep Learning (2006-)

Deep learning is the application of artificial neural networks with multiple hidden layers that form multiple levels of representations corresponding to different levels of abstraction

- New activation functions (ReLU, etc)
- New regularization techniques (dropout, etc)
- New learning techniques (SGD, Adam)
- Unsupervised representation learning and autoencoders
- Convolutional neural networks (CNN)
- Recurrent neural networks (RNN)
- Generative adversarial networks (GAN)

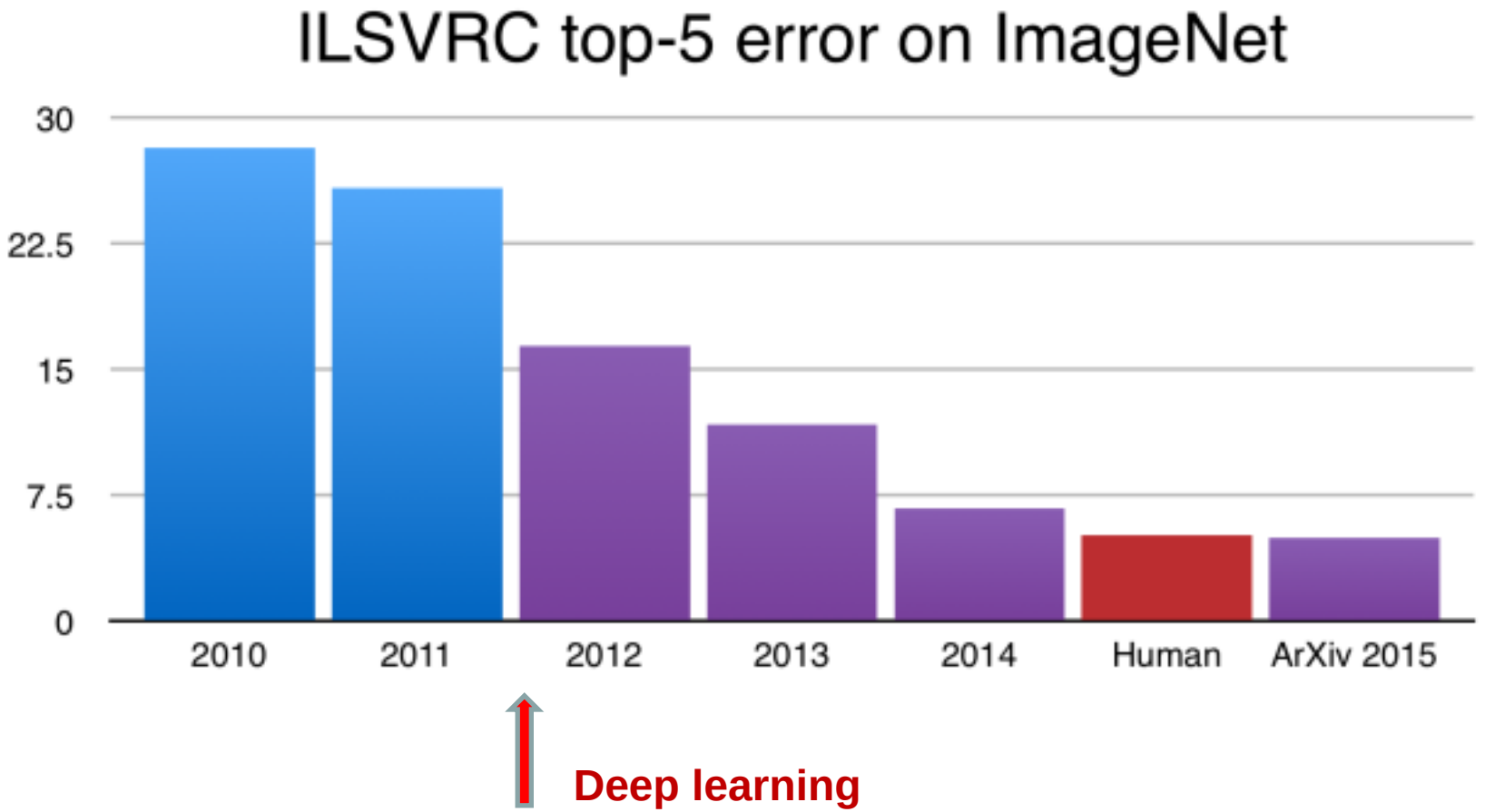


Jeffrey Hinton

Yoshua Begnio

Yann LeCun

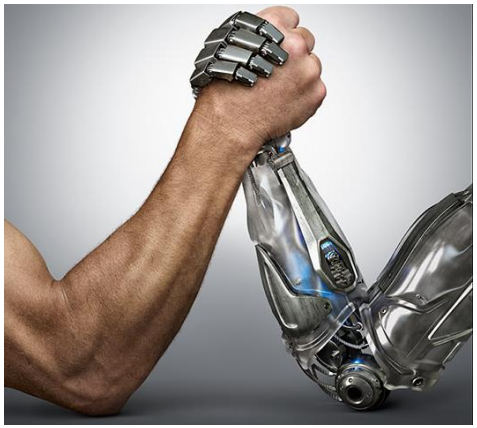
Evolution of Image Recognition Performance



http://houseofbots.com/news-detail/2575-4-deep-learning-vs-classical-machine-learning

Successful Applications of Deep Learning

- Image recognition
- Speech recognition
- Natural language processing
- Playing games (Go, etc)
- Self-driving cars
- Medical diagnostics
- etc



Deep learning beats humans

Revolution in the field of artificial intelligence

Big hype in media

Open Questions and Problems with Deep Learning
(Why are we still far from general (strong) AI?)

- Causality *(the number of storks vs the birth rate*)
- Extrapolation (*the problem of AD of QSAR models*)
  - DL models can easily be fooled
- No strong mathematical foundation (*contradicts statistical learning theory*)
- DL requires a lot of labelled data
- No neurophysiological evidence for backpropagation
- etc

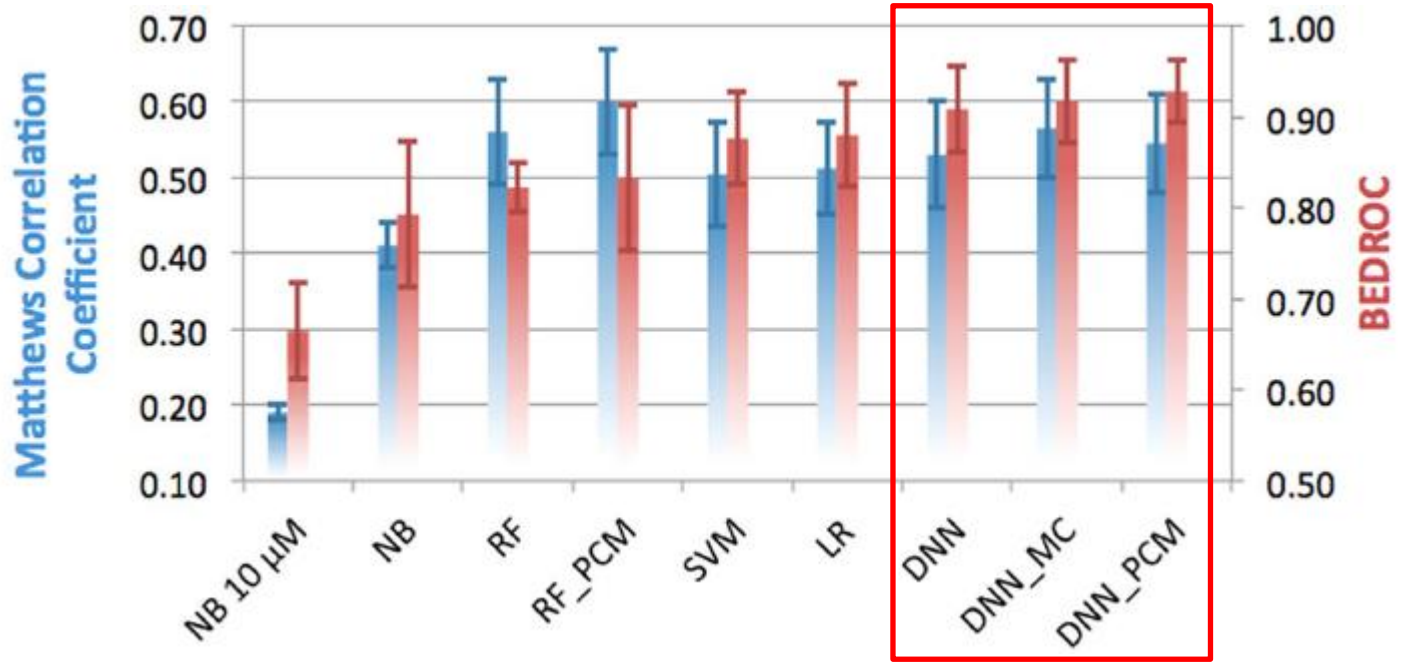
Lenselink et al. J Cheminform (2017) 9:45
DOI 10.1186/s13321-017-0232-0

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set

Eelke B. Lenselink[1†], Niels ten Dijke[2], Brandon Bongers[1], George Papadatos[3,4], Herman W. T. van Vlijmen[1], Wojtek Kowalczyk[2], Adriaan P. IJzerman[1] and Gerard J. P. van Westen[1*†]



Random split validation

Creativity

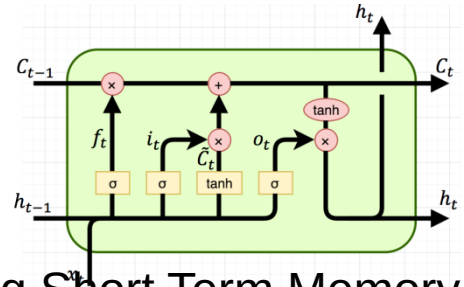
Creativity is the ability to create new objects based on generative models

Neural networks can create?

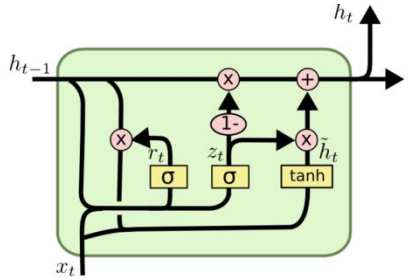
- New pictures (images)
- New music
- New posts in twitter
- New stories
- New poems
- New movie plots
- New dances
- etc
- …
- New molecules with desired properties

Creative Neural Networks

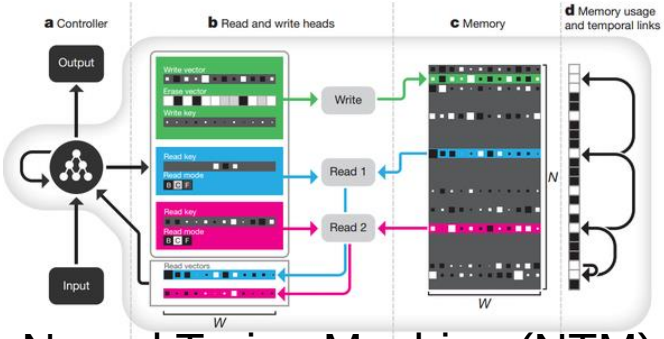
Recurrent Neural Networks
(repetitive running of neurons)



Long Short Term Memory (LSTM)

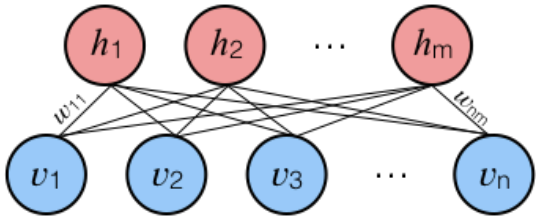


Gated Recurrent Units (GRU)

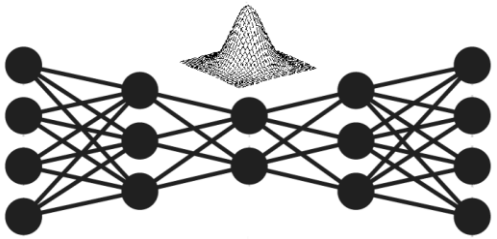


Neural Turing Machine (NTM)

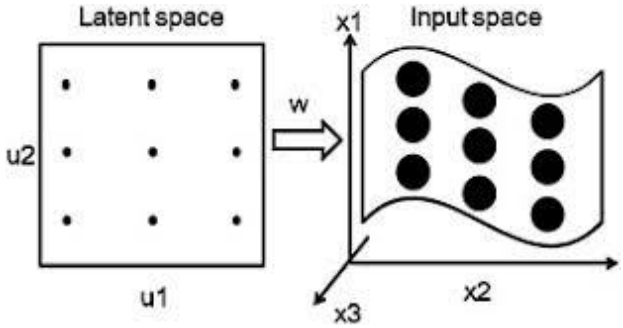
Energy-Based Neural Networks
(sampling form statistical distributions)



Restricted Boltzmann Machine (RBM)



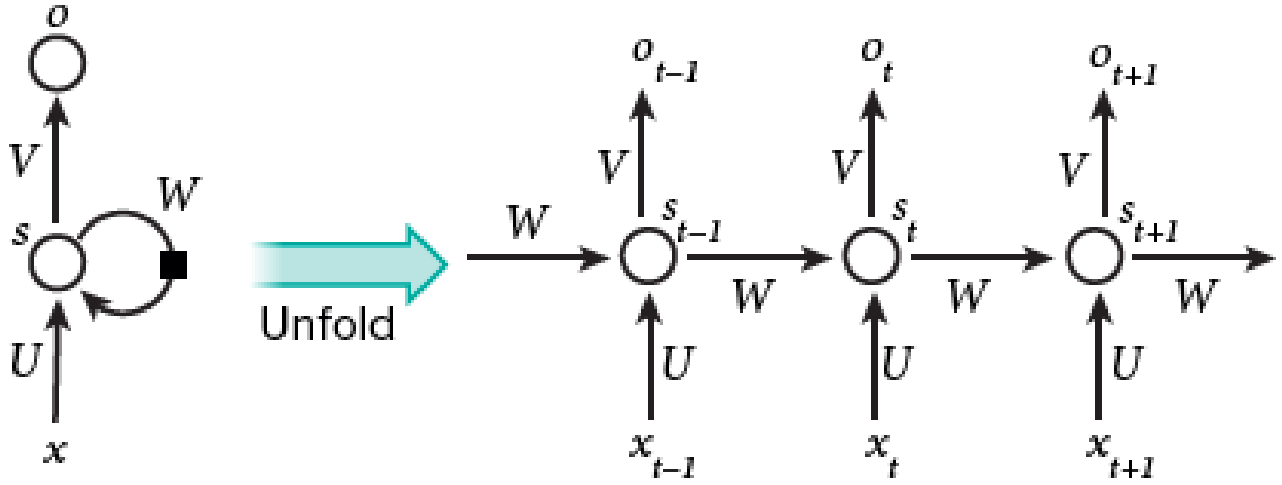
Variational Autoencoder (VAE)

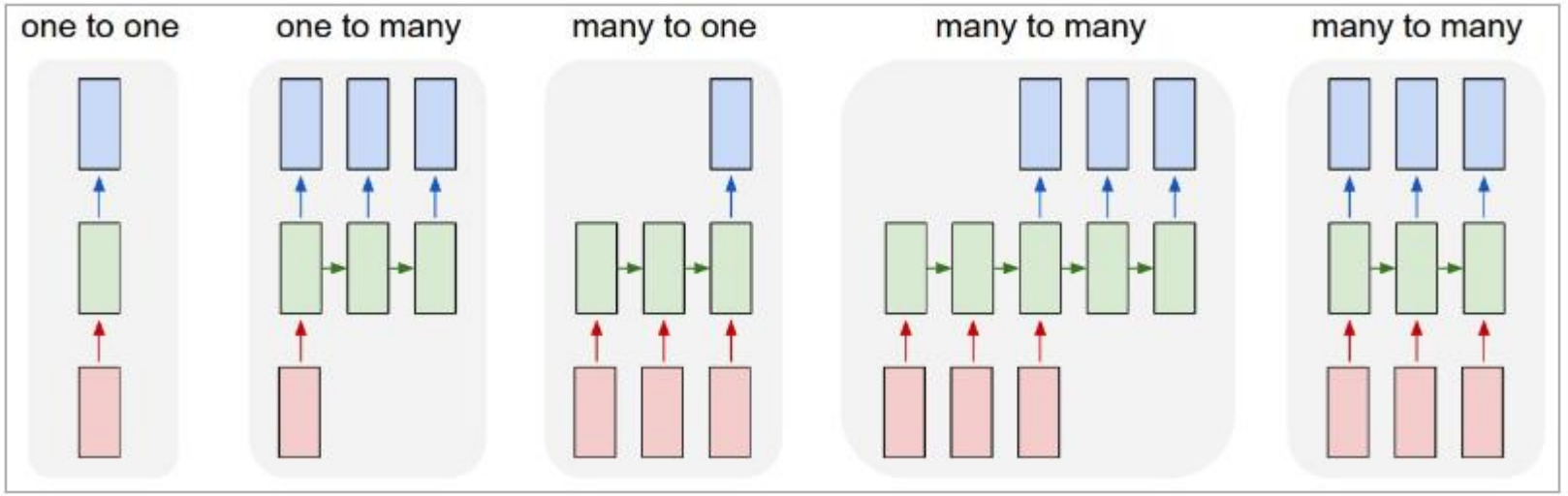


Generative Topographic Mapping (GTM)

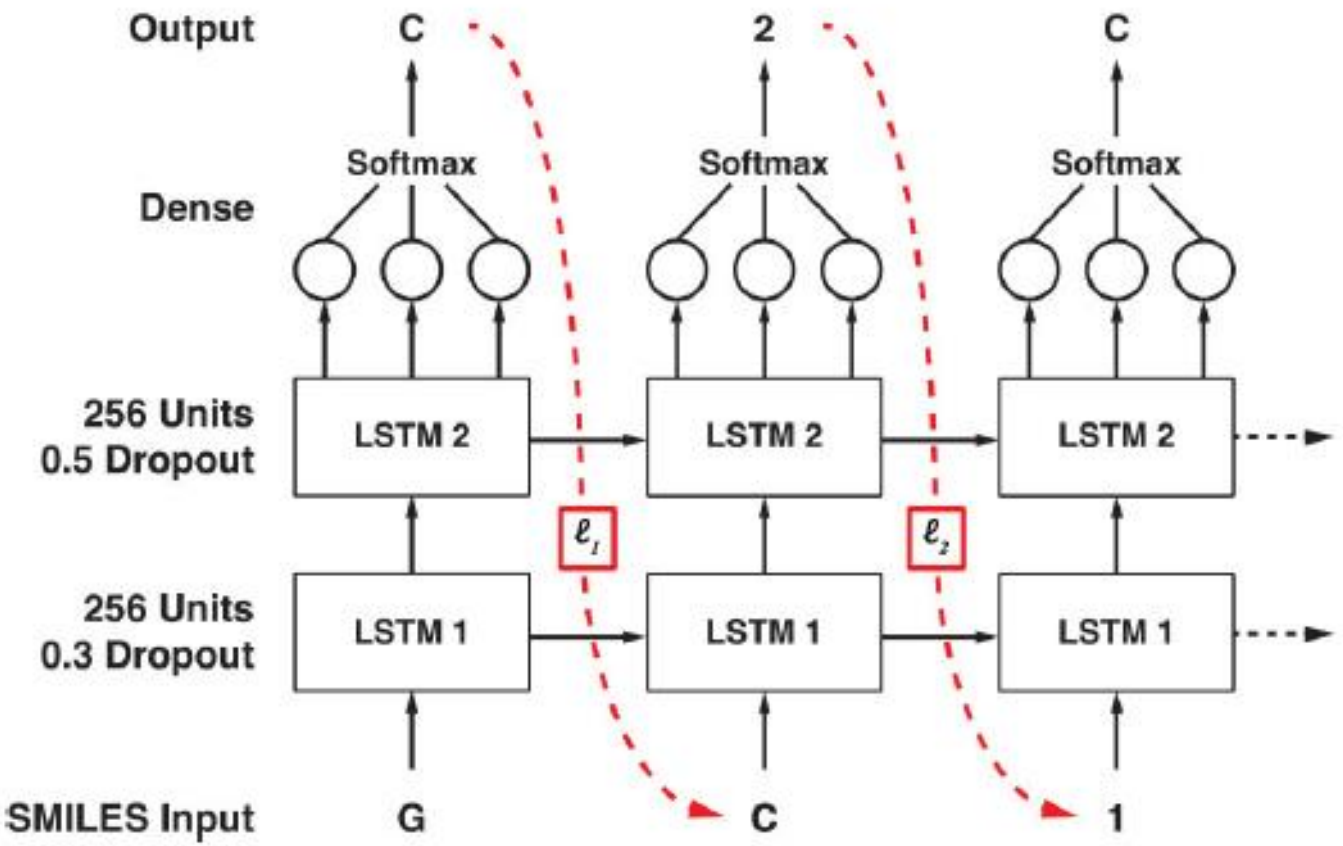
Recurrent Neural Networks (RNN)

Unfolding RNN in Time





Processing SMILES by Recurrent Neural Network



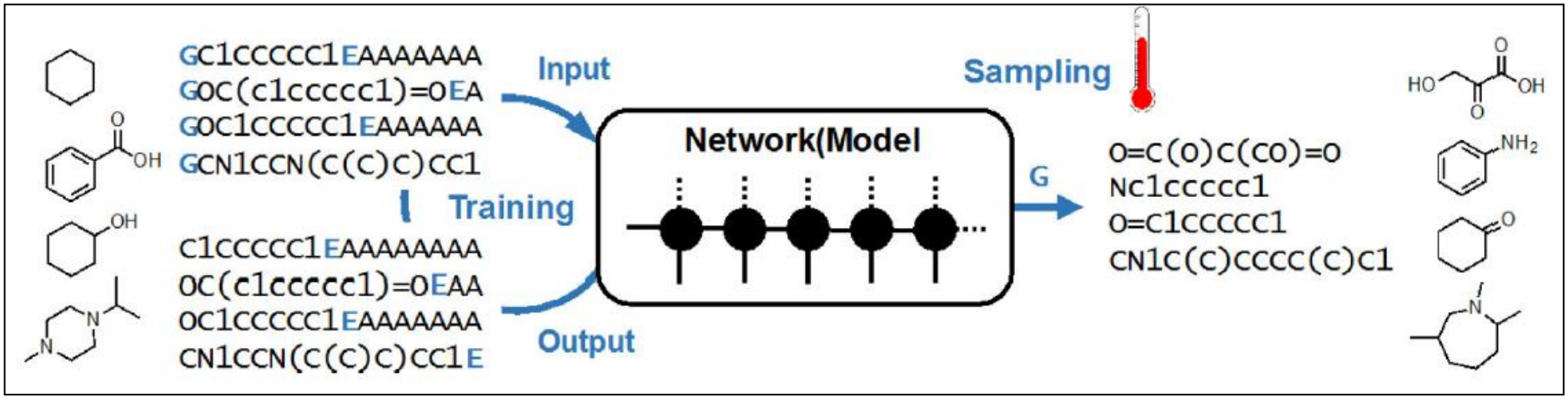
The network is producing SMILES strings, character by character (or token by token). During training, the network is learned predicts next character in current SMILES.

A. Gupta et al., *Mol. Inf.*, **2017**, *36*, 1700111

DOI: 10.1002/minf.201700111

Generative Recurrent Networks for *De Novo* Drug Design

Anvita Gupta,[a, b] Alex T. Müller,[a] Berend J. H. Huisman,[a] Jens A. Fuchs,[a] Petra Schneider,[a, c] and Gisbert Schneider*[a]



1. RNN Is trained to generate valid SMILES strings with high accuracy
2. The model is fine-tuned to specific ligand subsets with certain biological activity
3. The model can perform fragment-based drug discovery by growing molecules starting from a known active fragment

A. Gupta et al., *Mol. Inf.*, **2017**, *36*, 1700111

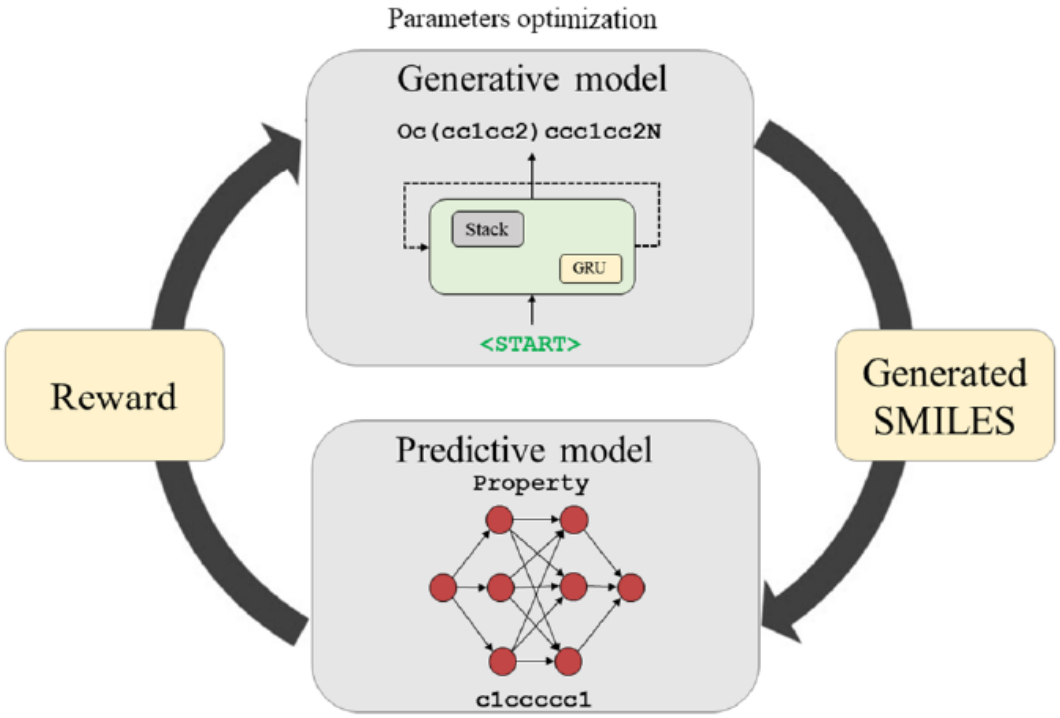
Reinforcement Learning (RL)

Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize reward (or minimize penalty)

- For successfully accomplished missions, the agent will be rewarded, and his experience will be learned by another agents
- If he will be killed, the next agents will not repeat his mistakes
- Agent 008 will be smarter

Deep Reinforcement Learning for *De-Novo* Drug Design

Mariya Popova[1,2,3], Olexander Isayev[1*], Alexander Tropsha[1*]

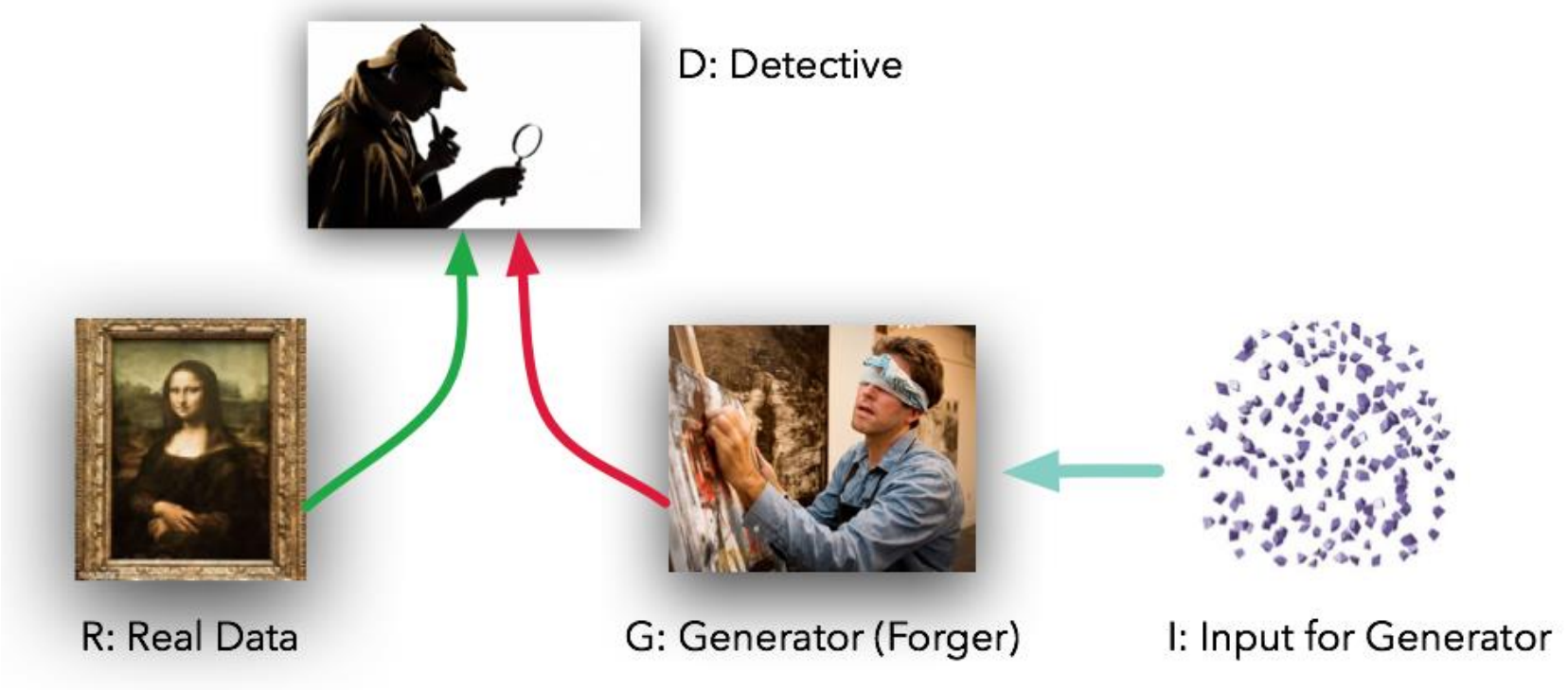


- **Mission** – build SMILES strings character-by character
- **Success** – SMILES corresponds to an active molecule

arXiv:1711.10907

Generative Adversarial Networks (GANs)

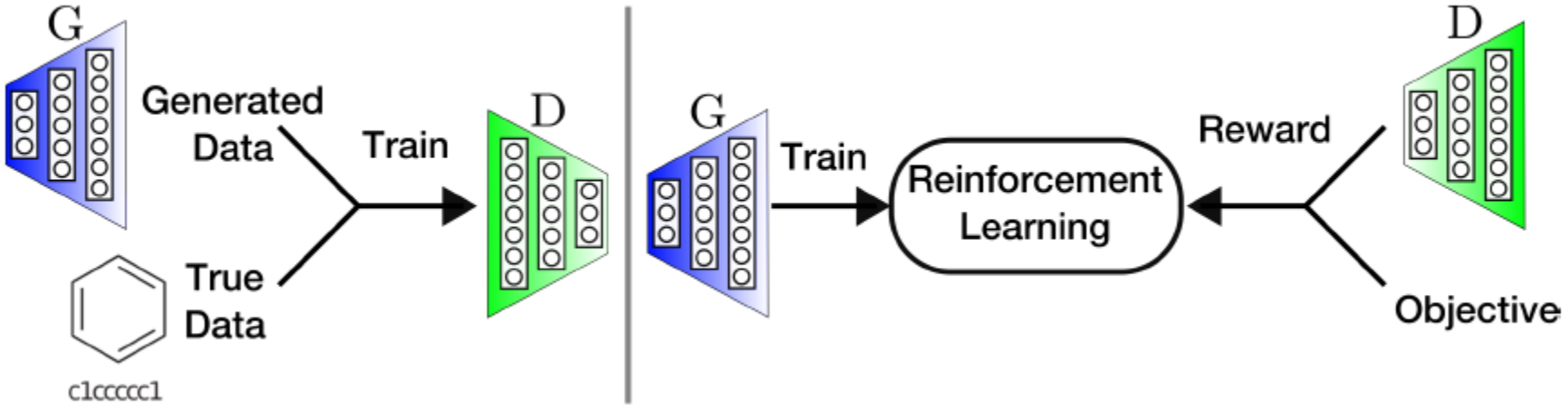
Generative adversarial networks (GANs) are a class of artificial intelligence algorithms used in unsupervised learning, implemented by a system of two neural networks contesting with each other (*Wikipedia)*



- Generator (forger) tries to fool discriminator (detective) with fake objects (pictures)
- Discriminator (detective) tries to improve its ability to detect fake objects (pictures)

https://medium.com/@devnag/generative-adversarial-networks-gans-in-50-lines-of-code-pytorch-e81b79659e3f

Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models

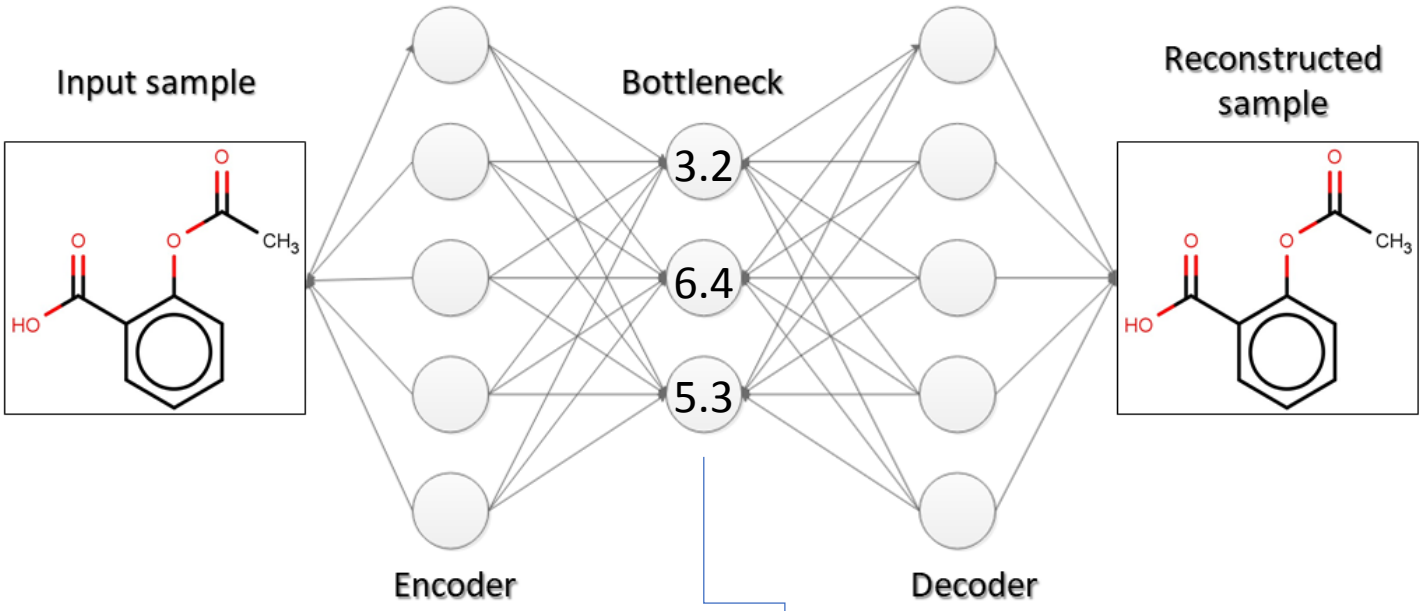


The generator G is trained to maximize two rewards at the same time:
(1) one that improves the activity of molecules
(2) another one that tries to mimic real structures by fooling the discriminator D

D – convolutional neural network (CNN)
G – long short term memory recurrent neural network (LSTM)

Guimaraes et al. arXiv:1705.10843

Autoencoder



- Is an artificial neural network
- Trained to accurately reconstruct its input object

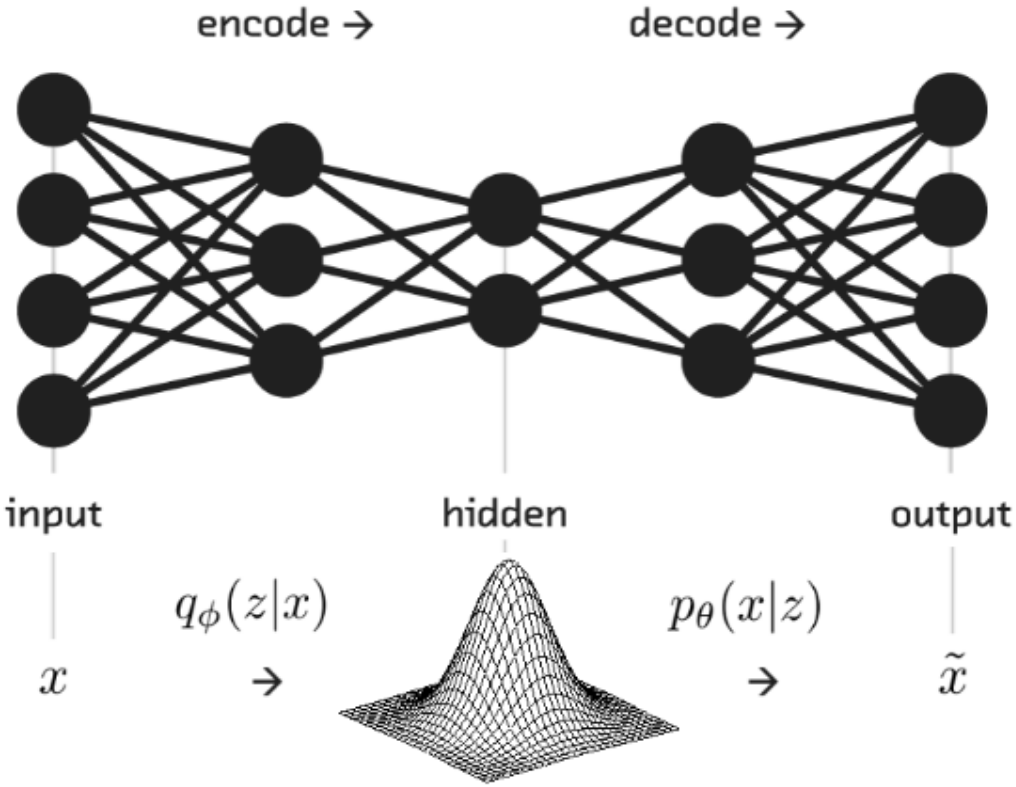
Latent vector

3.2
6.4
5.3

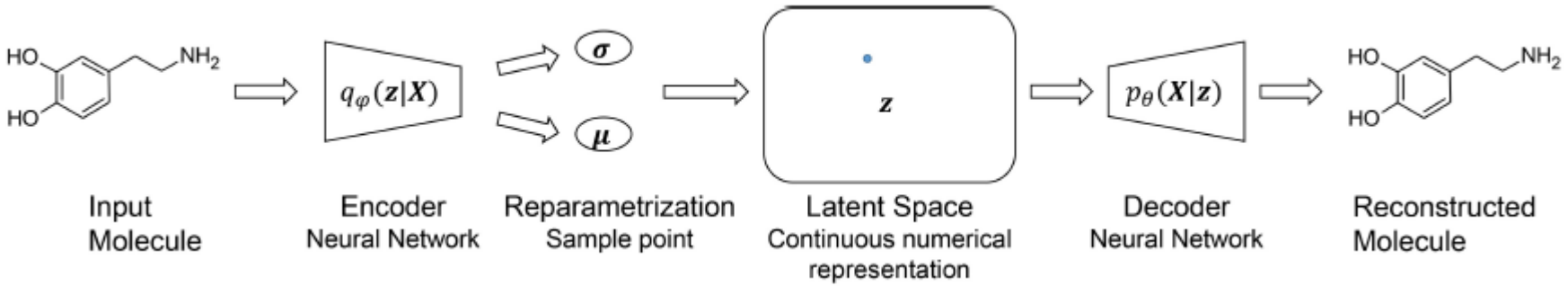
Encoder transforms chemical structure into a ***Latent vector*** - unique *reversible* representation in real numbers

Variational Autoencoder (VAE)

R.Feyman:
“What I cannot create, I do not understand”



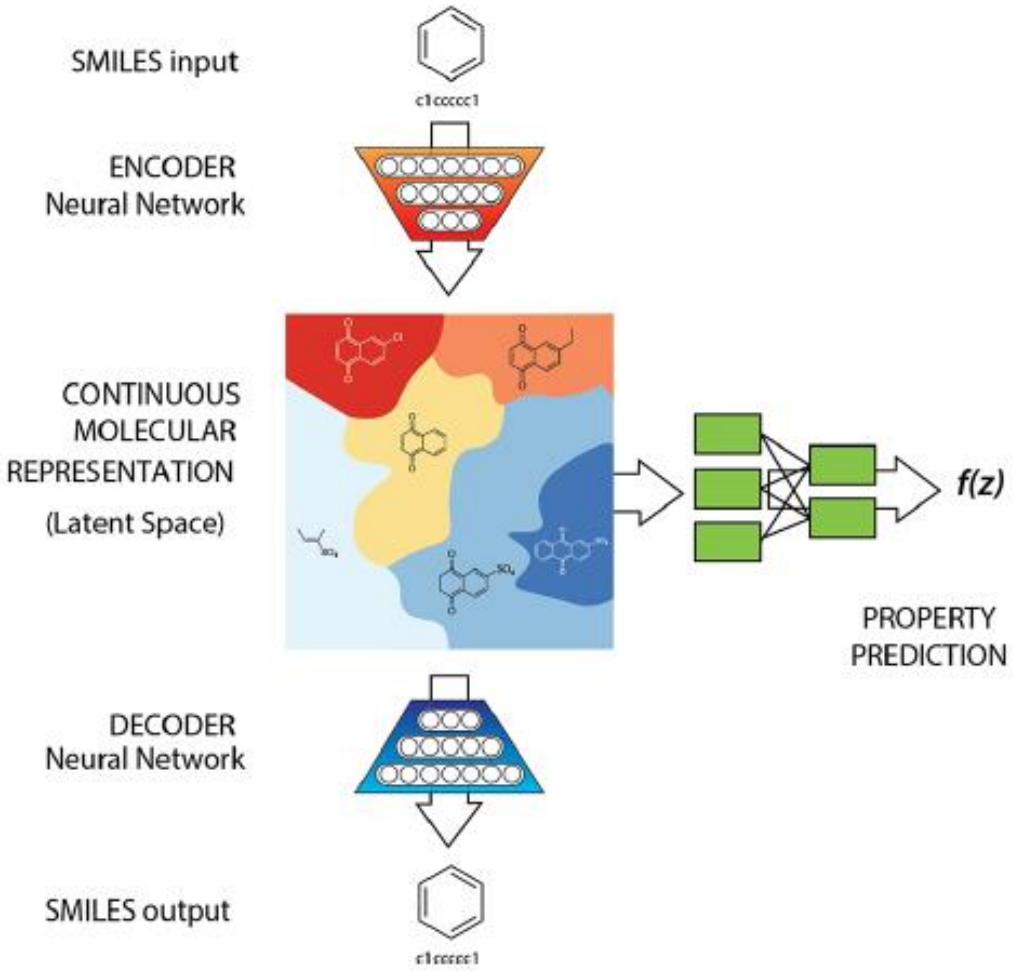
arXiv:1312.6114

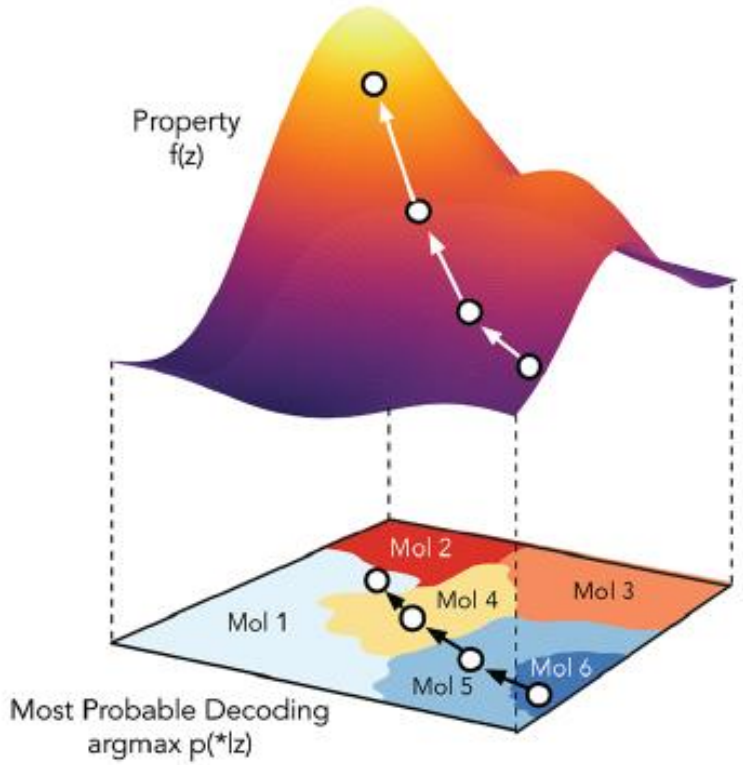


T. Blaschke et al. *Mol. Inf.*, **2017**, *36*, 1700123

Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

Rafael Gómez-Bombarelli,[†,#] Jennifer N. Wei,[‡,#] David Duvenaud,[¶,#] José Miguel Hernández-Lobato,[§,#] Benjamín Sánchez-Lengeling,[‡] Dennis Sheberla,[‡] Jorge Aguilera-Iparraguirre,[†] Timothy D. Hirzel,[†] Ryan P. Adams,[∇,∥] and Alán Aspuru-Guzik*[,‡,⊥]



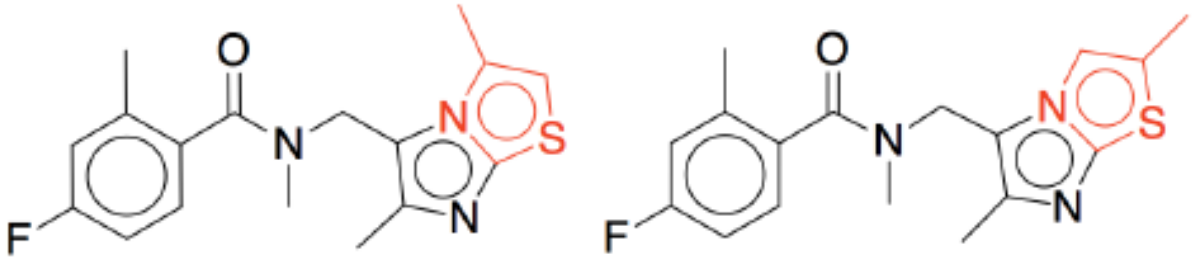


ACS Cent. Sci., **2018**, *4*(*2*), 268-276

Problems with SMILES-Based Molecule Generation

1. Correct SMILES should be generated, so their grammar should be learned from data*
2. SMILES are not unique for a molecule, but their latent representations should be the same
3. SMILES cannot be formed for generalized structures
4. SMILES have bad neighborhood behavior

*fixed with GrammarVAE (arXiv:1703.01925)



Canonical SMILES:

Cc1cn2c(CN(C)C(=O)c3ccc(F)cc3C)c(C)nc2s1

Cc1cc(F)ccc1C(=O)N(C)Cc1c(C)nc2scc(C)n12

Graph-Based Generation of Molecules by Neural Networks

1. M. Simonovsky et al., “GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders”, *arXiv*:1802.03480, 9 Feb **2018**
2. W. Jin et al., “Junction Tree Variational Autoencoder for Molecular Graph Generation”, *arXiv*:1802.04364, 19 Feb **2018**
3. Y. Li et al., “Learning Deep Generative Models of Graphs”, *arXiv*:1803:03324, 8 Mar **2018**
4. N. De Cao et al. “MolGAN: An implicit generative model for small molecular graphs”, *arXiv*:1805.11973, 30 May **2018**
5. J. You et al. “Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation”, *arXiv*:1806.02473, 7 Jun **2018**

- Molecular graphs are generated at once
- GAN, VAE and RL frameworks are combined with “isomorphism-aware” approaches

Two Kinds of Creativity

Disruptive creativity

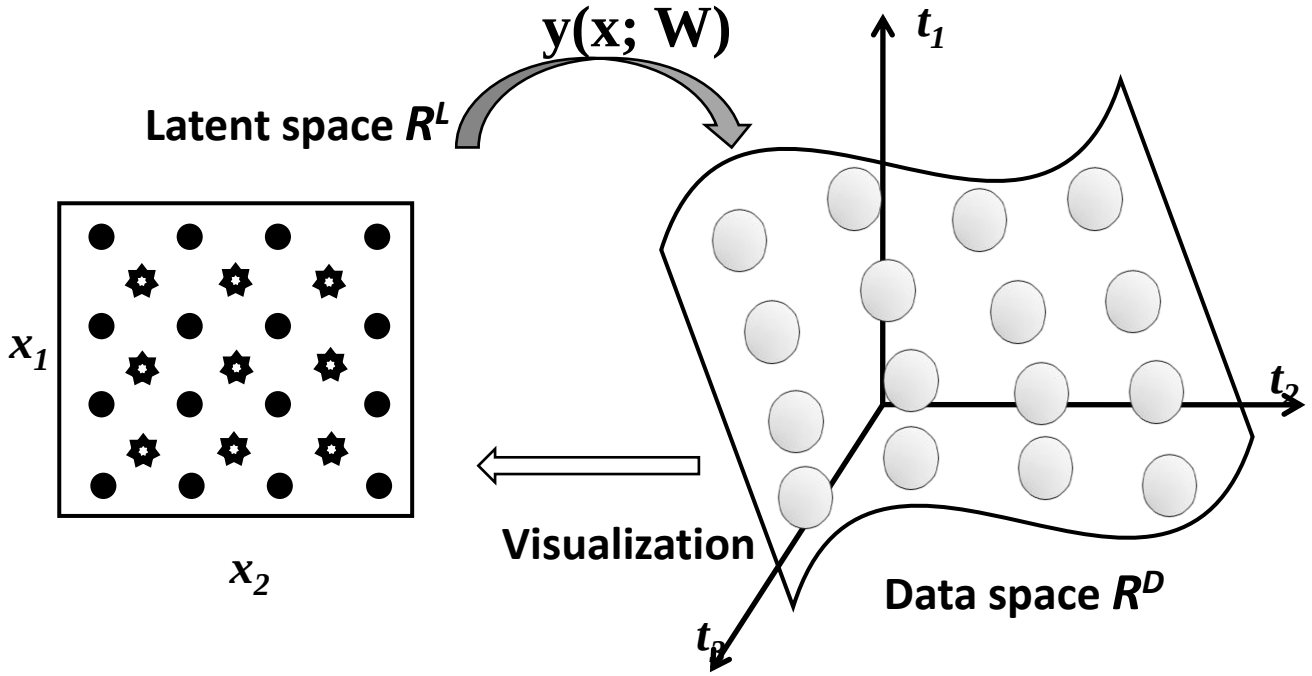
New ideas completely replace the old ones

Conservative creativity

New ideas are added and combined with the old ones proved to be useful

Generative Topographic Mapping (GTM)

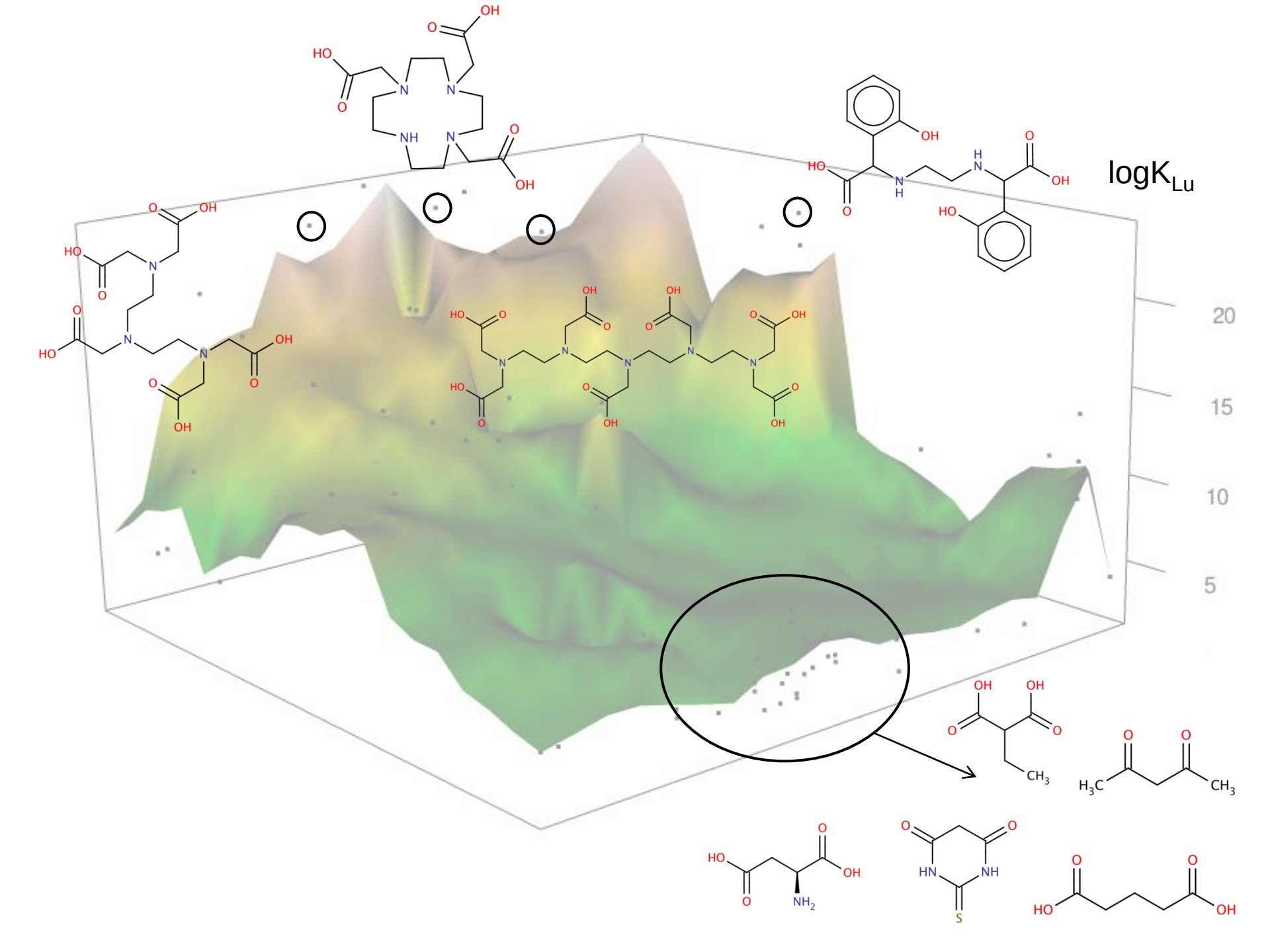
Christopher M. Bishop



• GTM relates the 2D or 3D latent space with a manifold embedded in the high-dimensional data space.

• The visualization plot is obtained by projecting the data points onto the manifold and then unfolding it.

C.M.Bishop, M.Svensen, C.K.I.Williams, *Neural Computations*, **1998**, *10*, 215-234


logK$_{Lu}$
20
15
10
5

GTM – Swiss Army Knife of Chemoinformatics

Comparison of chemical databases

Visualization and cartography of chemical space

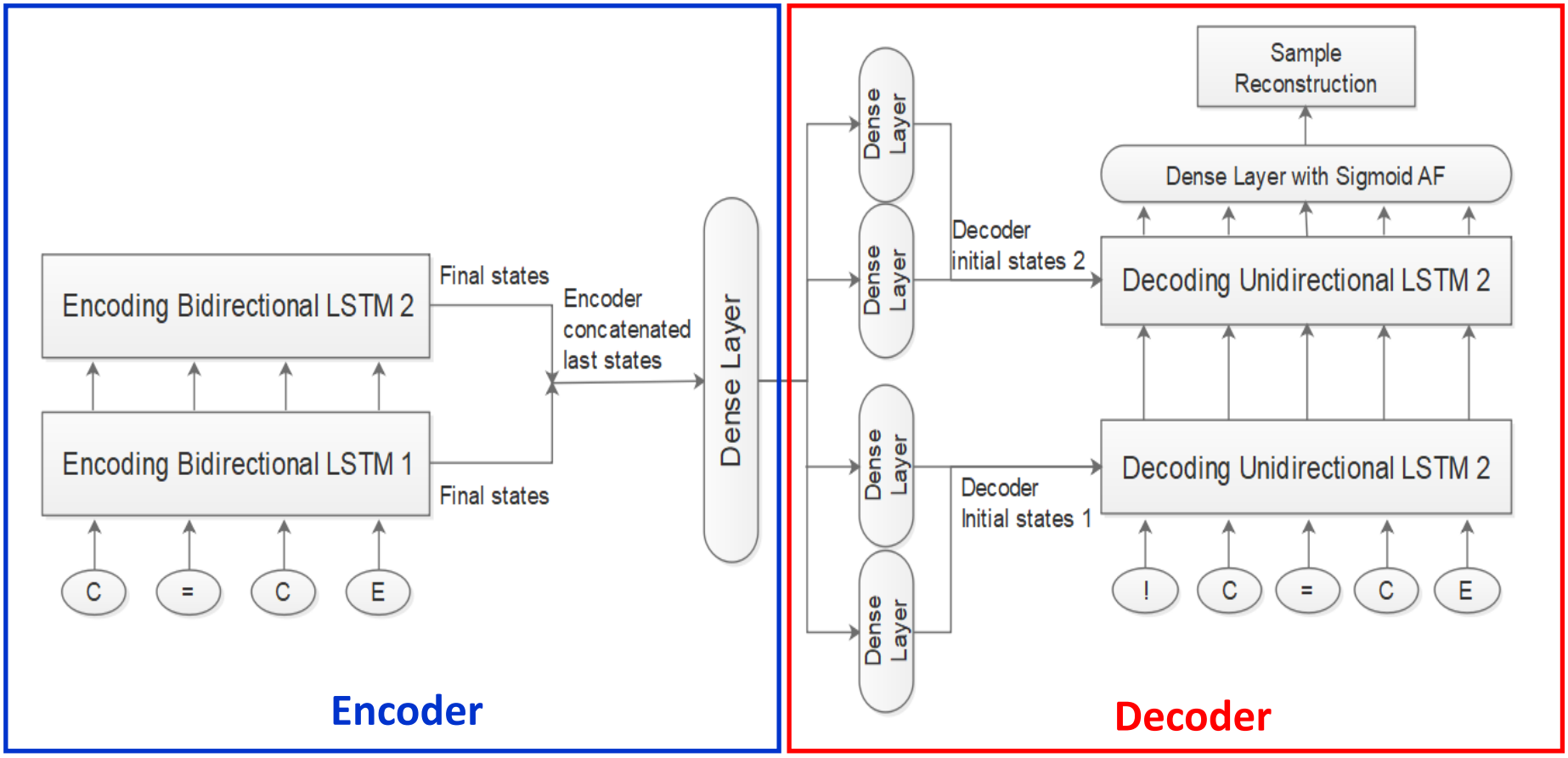
Classification models

Regression (QSAR) models

Virtual screening

Inverse QSAR

Implementation of Autoencoder Architecture



- **Encoder** consists of two bidirectional LSTM layers followed by a dense layer
- **Decoder** consists of two unidirectional LSTM layers and several dense layers

- Trained on 1.2 M ChEMBL compounds
- Reconstruction accuracy: **95.81%** on test set (300 K)

Generation of Focused Library for A2a Receptor

Adenosine A2a receptor activity dataset

Structures

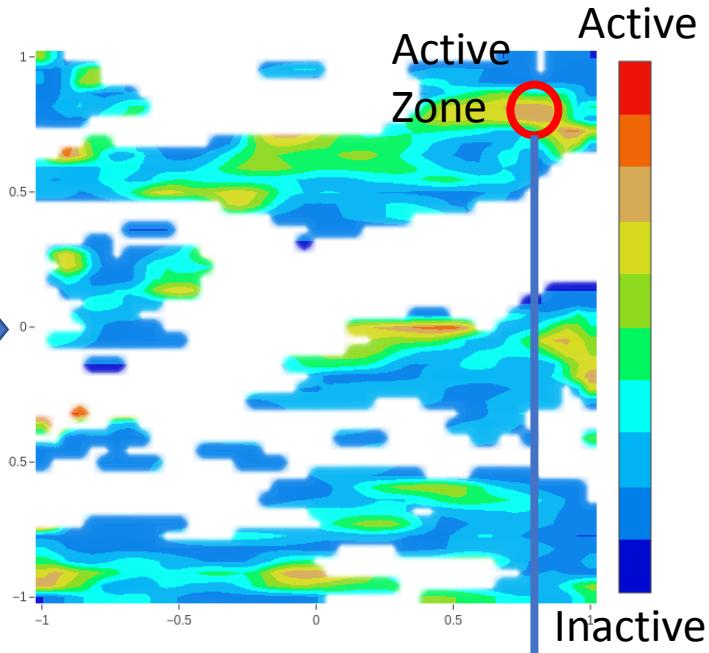
Active 3618

Inactive 1303

SMILES

Trained ENCODER

Latent Vectors



"Active" Latent Vectors

Trained DECODER

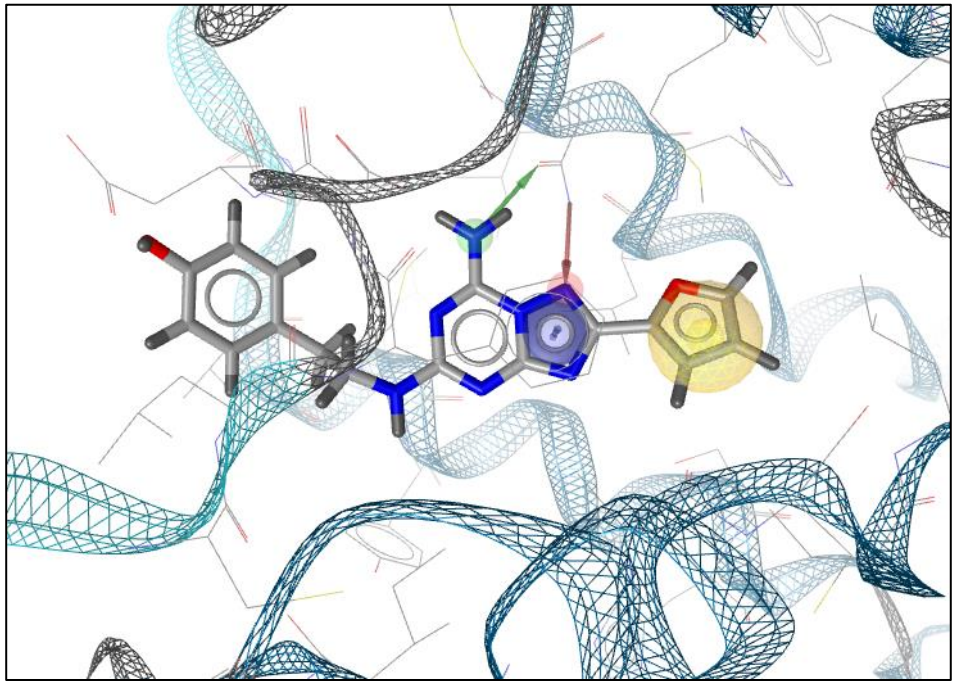
SMILES

Focused library:
27034 synthesizable*
molecular structures

*synthetic accessibility estimated by P.Etrl (2009) model

Validation Using Pharmacophore Model

- Structure-based pharmacophore model based on PDB structures of A2a receptor with co-crystallized ligands has been developed
- This model was applied to (*i*) generated molecules and (*ii*) a subset randomly extracted from ChEMBL database



Results of virtual screening using pharmacophore:

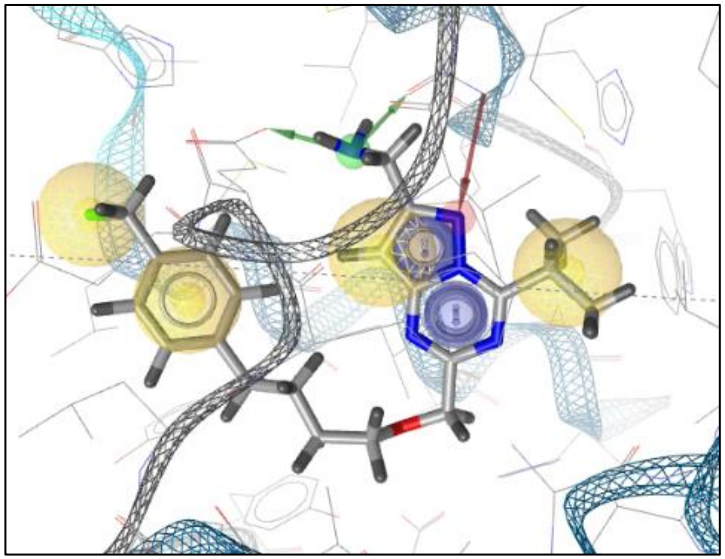
Generated library, hits rate, %	CheMBL23, hits rate, %
16.05 (i)	5.37 (ii)

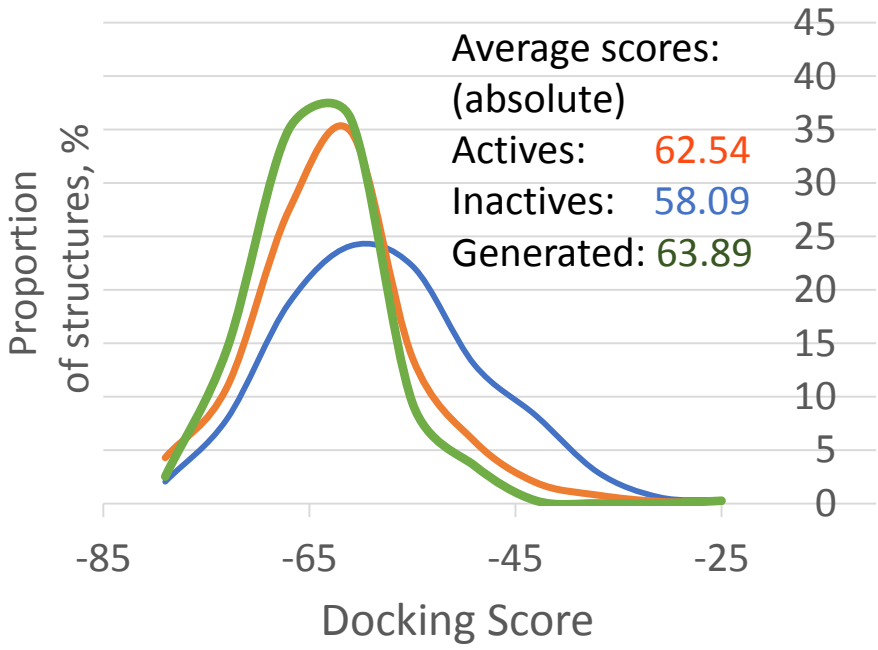
De novo generated library is 3 times more likely to comply to the developed pharmacophore model than ChEMBL23 random baseline

Validation Using Ligand-Protein Docking

Docking was performed on 2YDO PDB x-ray structure of A2a receptor using the S4MPLE tool

Top scored docking pose of one of generated molecules

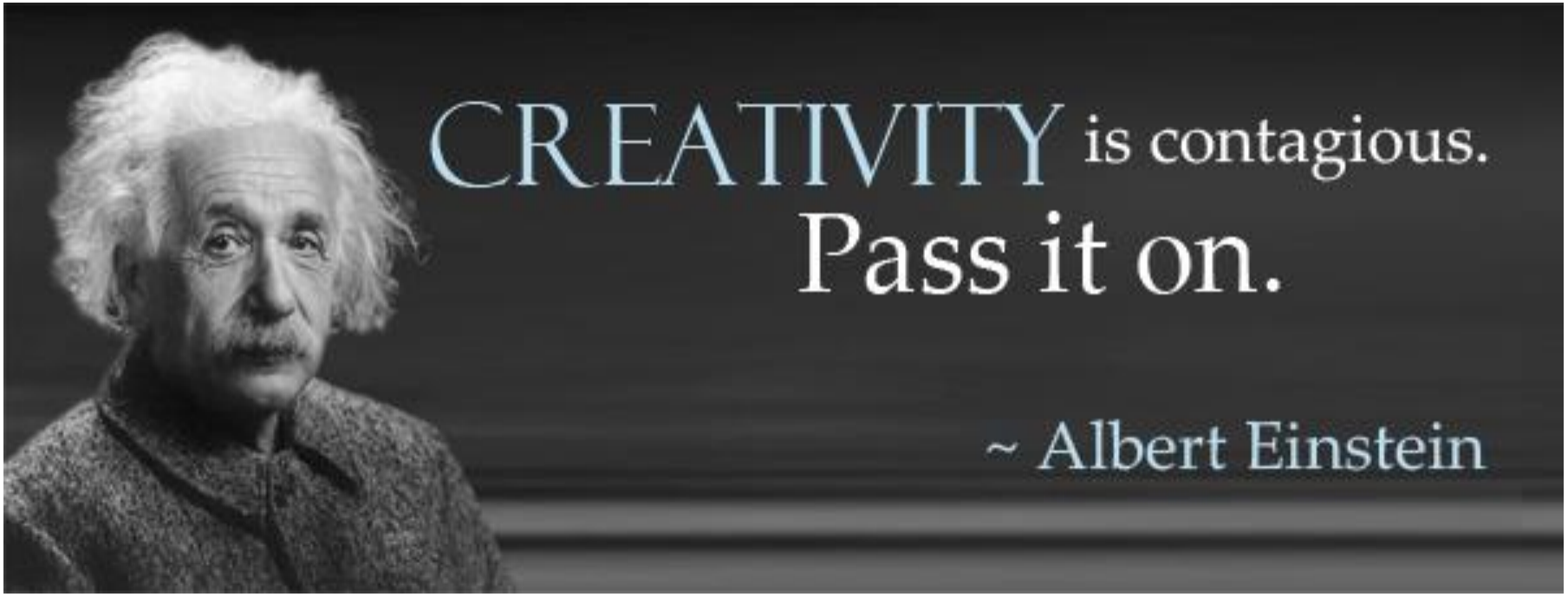




- Distribution of the docking scores of generated structures closely follows distribution of real active molecules
- Average score of generated molecules is even higher than for real actives

Conclusions

- It is creativity that makes neural networks great again in chemoinformatics
- Conjunction of DL with GTM results in a novel de novo molecular design method that combines the creative character of RNN with unique data visualization capabilities of GTM


CREATIVITY is contagious.
Pass it on.
~ Albert Einstein

Acknowledgement

Boris Sattarov

Alexandre Varnek

Dragos Horvath

Gilles Marcou