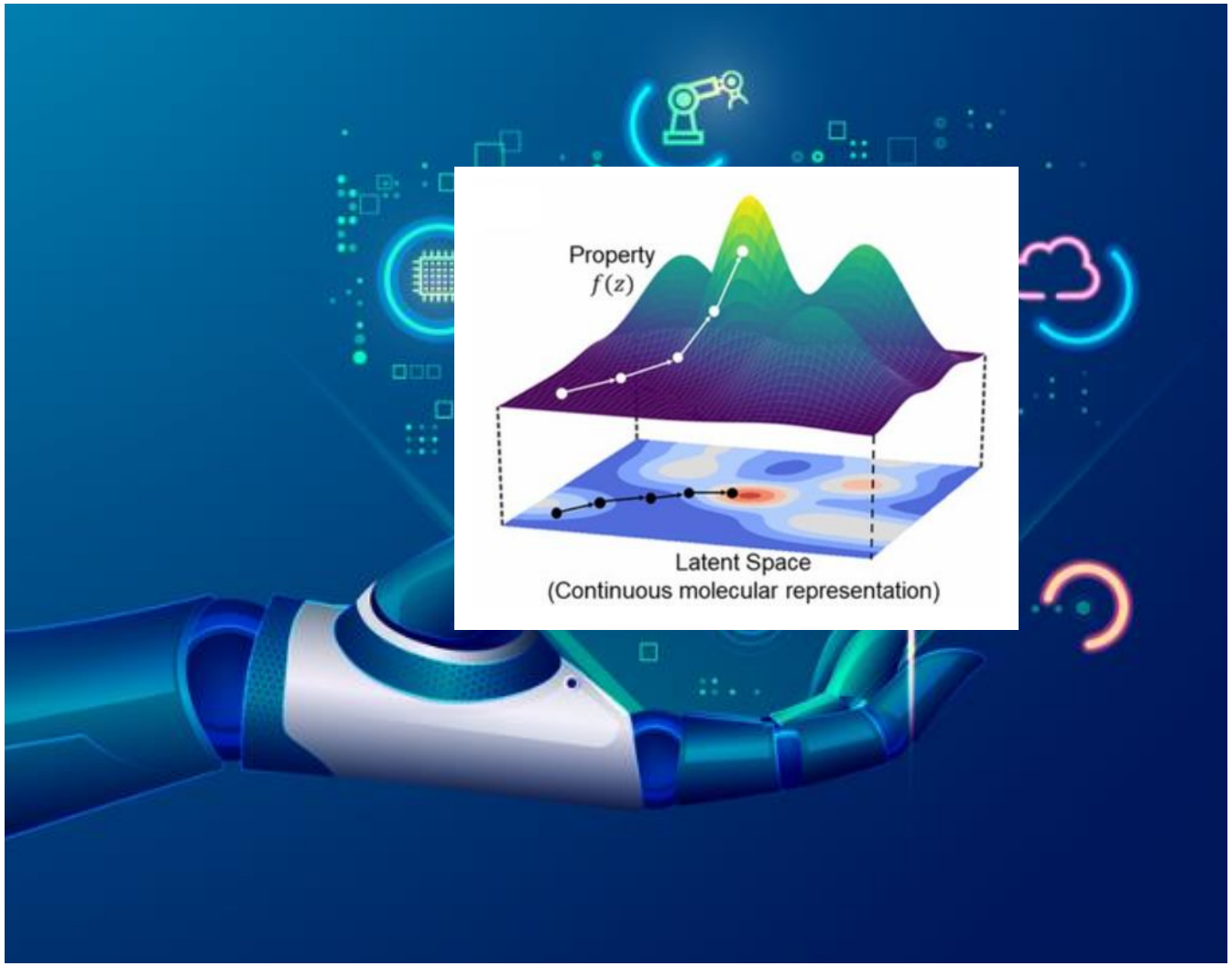
RAVEN – AI generated molecules in drug like space

Dr. Mark Griffiths
Principal Data Scientist I
Machine Learning & Artificial Intelligence (RAIDERS)
Department of Data Solutions
R&D Information Research

AbbVie Deutschland GmbH & Co. KG



Pistoia Alliance Spring Conference, 24th April, London

abbvie

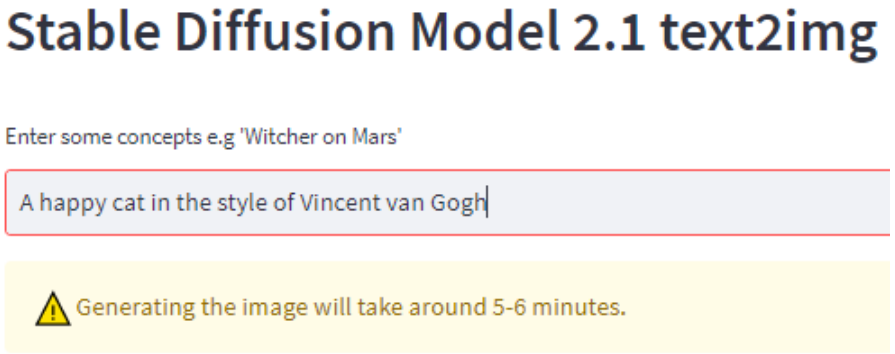
RAIDERS

RAVEN – AI generated molecules in drug like space

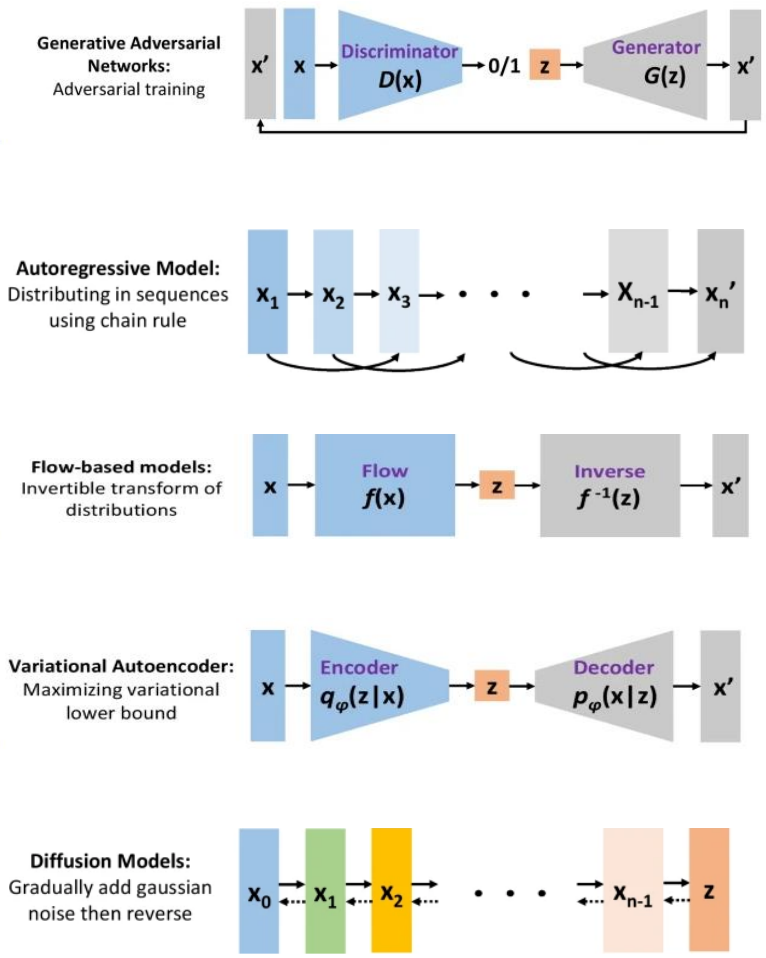
Relative **A**ttention based **V**ariational auto-**EN**coder

.

Generative AI refers to a group of machine learning models and algorithms such as generative adversarial networks (GANs), variational autoencoders (VAEs), autoregressive models, flow-based generative models, and diffusion models that have been applied successfully to generate new content in the form of images, text, audio, and more.



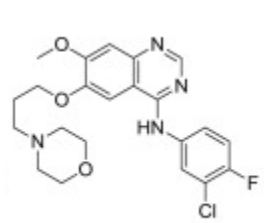
RAVEN – AI generated molecules in drug like space

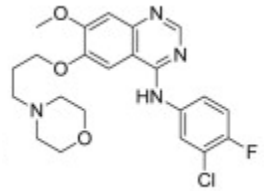


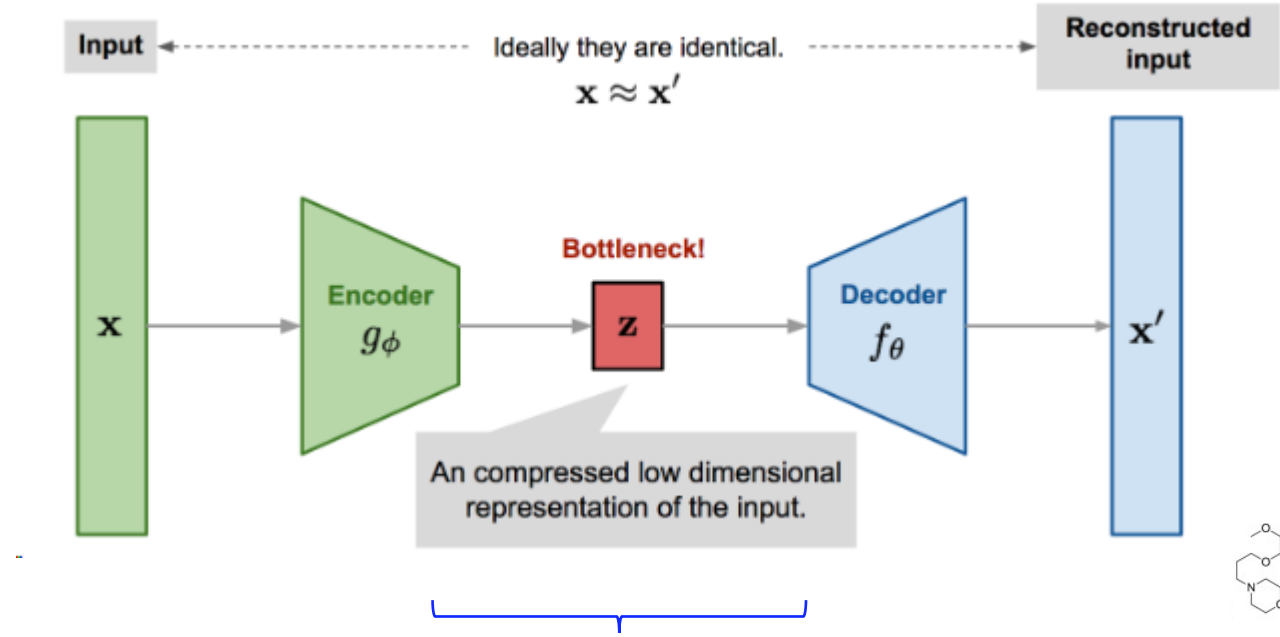
- GANs consist of two neural networks: the generator and the discriminator. The generator generates new outputs, while the discriminator evaluates the generator's outputs and identifies them as either real or generated. The generator is trained to improve its outputs by trying to fool the discriminator into believing that they are real and not generated.
- Autoregressive models are another type of generative AI capable of generating sequences of data, such as text, music, or speech. The model learns to determine the probability of the next elements in the sequence based on the previous ones.
- A flow-based generative model is constructed by a sequence of invertible transformations.
- A variational autoencoder consists of an encoder and a decoder, acts as an identity function, and learns to output whatever is inputted. By forcing the model to become an identity function it stores all the input data's relevant features in a compressed representation called a latent space which can be used to generate new output.
- A diffusion model works by gradually adding Gaussian noise to the original data in the forward diffusion process and then learning to remove the noise in the reverse diffusion process.

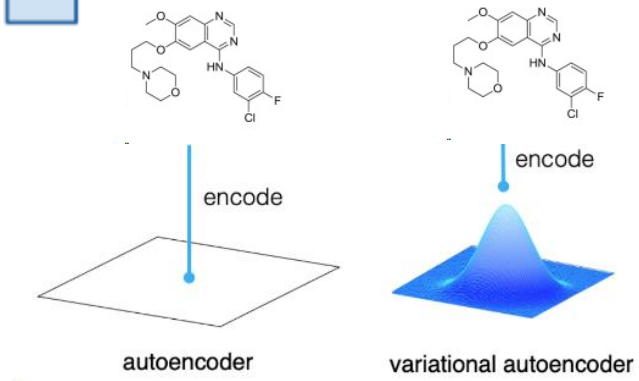
RAVEN – AI generated molecules in drug like space

Autoencoder, by forcing the model to become an identity function it stores all the input data's relevant features in a compressed representation called a latent space which can be used to generate new output.









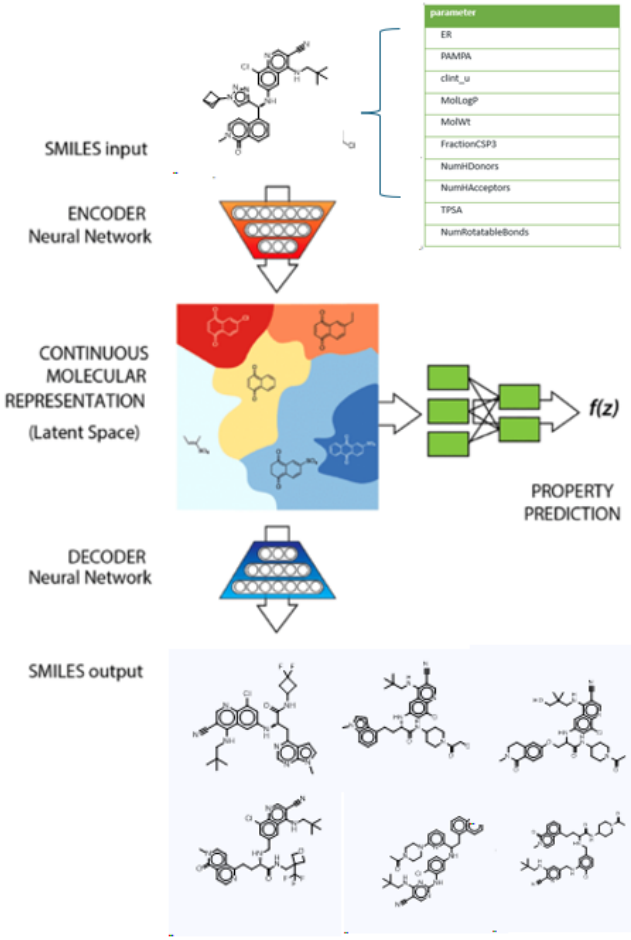
https://arxiv.org/abs/1312.6114

abbvie

RAVEN – AI generated molecules in drug like space

Deep generative models have been applied successfully to generate text, images, video and music, in RAVEN we changed the paradigm from **predictions of properties** to **predictions of which molecules to create and test**.

- In house developed Variational Autoencoder (VAE) LLM model built using internal and external data.
- The model has 13 billion parameters. When given a prompt with a seed molecule (optional) and ranges of desired parameters it starts generating molecules.
- Used in internal projects to generate new molecules and to create analogues of lead compounds with improved properties.
- The molecules generated were so good that chemists agreed to synthesize some, and several new lead compound series were identified.
- Thus, we have the first AI (LLM) generated molecules sitting in vials at AbbVie.

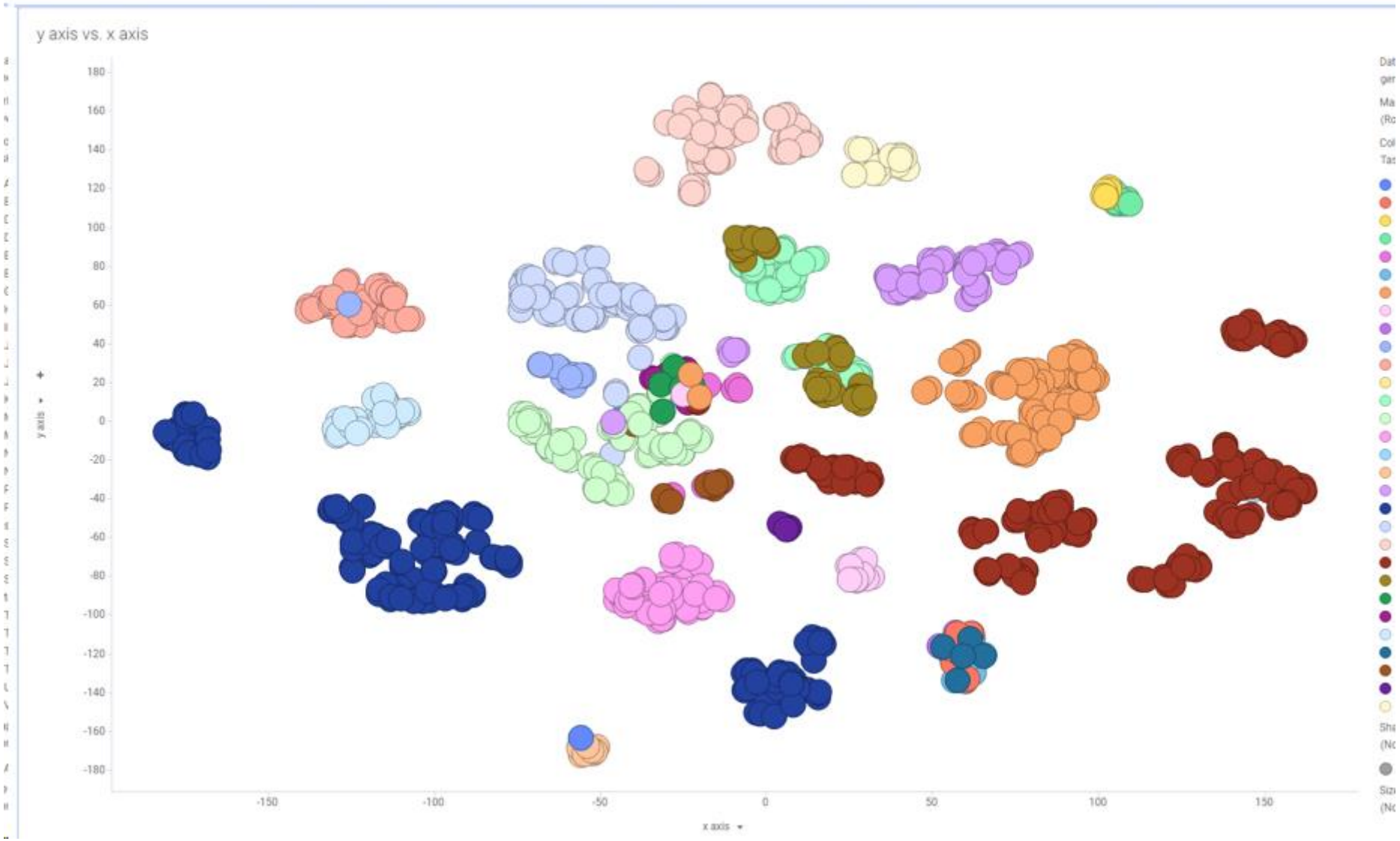


<pred_mol>clint_hu,clint_hum,clint_hr,1.7,clint_um,clint_ur,1.17,fu_m,mdc
,15.61,rb_fu,r_invivo_cu,rod_plasma_fu,so,vdss,vdss_m|1.5,457.87,0.33,1.0
,2.0,106.48
,3.0|O=C(O)C[C@H](Cc1ccccc1)Nc1nc2ccccc2cc1NS(=O)(=O)c1ccc2oc(=O)[nH]c2c1

“drug like space”

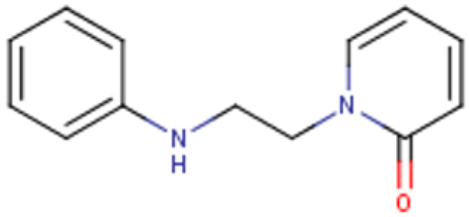
RAVEN – AI generated molecules in drug like space

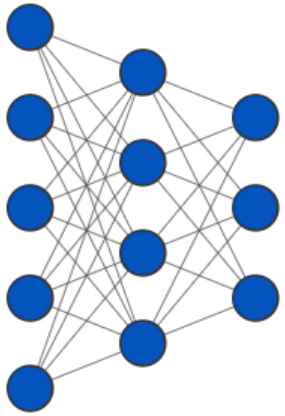
The chemical space for the recent RAVEN generated molecules consists of clusters which correspond to the projects / hint molecules used which makes sense.

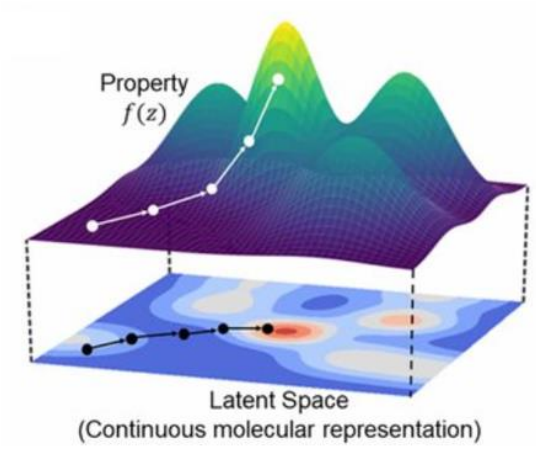


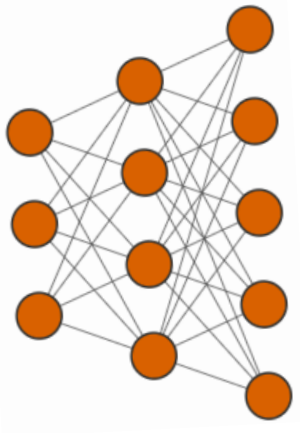
abbvie

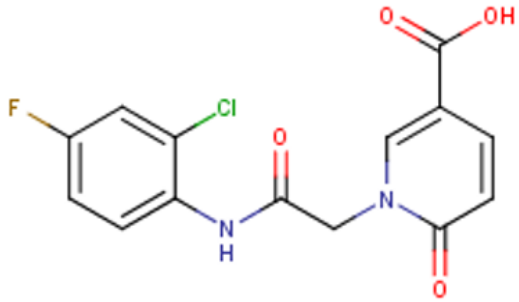
RAVEN – AI generated molecules in drug like space











Reinforcement Learning

abbvie

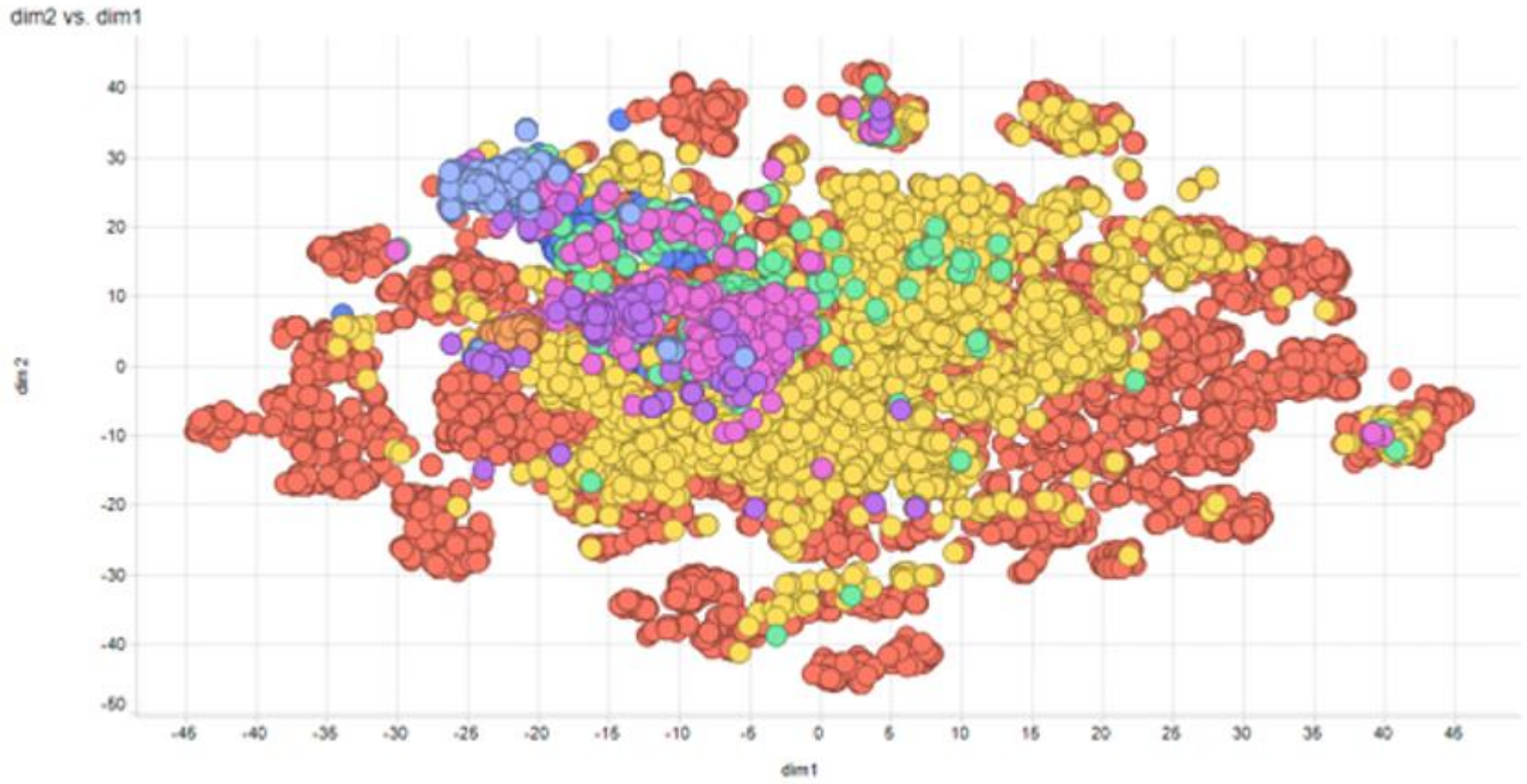
RAVEN – AI generated molecules in drug like space

Comparing RAVEN with Chem42 generated molecules using a Gilead molecule as hint.

The molecules are very different, the mw is lower for Chem42 (median 420 vs 640) as is the TPSA (93 vs 126).

RAVEN – AI generated molecules in drug like space

t-SNE scatter plot of Tanimoto distance between 4 generative methods including RAVEN for the same project, each method is producing a different chemical space.



abbvie

RAVEN – Conclusions

- We have created the RAVEN model, a generative language model based on VAE that is now available in the drug discovery toolbox at AbbVie.
- My experience is that you do not need to ask for 1000s of molecules asking for 50 or 100 is normally enough to get interesting ideas that can be followed up on or fed back into the model.
- The RAVEN model is being used in a variety of internal projects and has generated molecules chemists were prepared to have made.
- It proved useful that the Model can offer two modes of sampling. Using the expansion mode provides greater diversity and uniqueness in the molecules generated while using the exploitation mode leads to closer analogues to the hint molecule being generated.
- We are working on reinforcement learning and adding a local potency model (QSAR) to the workflow for some projects.
- We are very happy to see molecules generated by RAVEN being sent for synthesis.