

# A Clear Example of AI Value For Drug Discovery Has Arrived

Jan. 11, 2018 by Andrii Buvailo

With all the hot discussions (for instance, [here](#), [here](#), [here](#) and [here](#)) going on right now among medicinal chemists, pharmaceutical researchers, and data scientists as to what artificial intelligence (AI) means for the future of drug discovery, the life science world has divided into “AI-believers”, “AI-atheists”, and “AI-agnostics”.

It is useless to repeat what has been many times said about successes of AI in areas like natural language processing, image processing, pattern recognition and self-driving cars ([here is the summary](#)), but few of us knew if those sort of results (or any meaningful results at all) could possibly be achieved with such complex systems as biological organisms... Finally, however, a hint of hope arrived.

Just came across a fresh publication in *Molecular Informatics* which pushes the limits in understanding the kind of things that AI can really do already today for the medicinal chemistry and drug discovery. And it sounds cool indeed.

In their paper titled De Novo Design Of Bioactive Small Molecules By Artificial Intelligence, authors explained how they had trained AI-model, gained new bioactive molecules out of it, and then tested the molecules against targets in hybrid reporter gene assays to confirm 1 selective nM activity lead and a back-up, both new to the chemical record.

A brief summary of how they did it was provided in this [Linkedin post](#) by Alfred Ajami, I am just copying the text here:

- 1) Train RNN/LSTM on SMILES of 500k bioactives (< 1uM) from ChEMBL22.
- 2) Fine tune by transfer learning to enable de novo generation of structures: use 25 fatty acid mimetics with known agonistic activity on retinoid X receptors (RXR) and/or peroxisome proliferator-activated receptors (PPAR)."
- 3) From fine-tuned AI model, pick 1000 SMILES, by fragment growing from the minimalist start fragment “-COOH” rejecting structures identical to any in the training set.

- 
- 4) a) Rank de novo designs for predicted effects on RXRs and PPARs using target prediction methods (SPiDER), shape and charge descriptors to determine the similarity of the designed compounds to known bioactive ligands; b) merge the individual screening lists to obtain a final set of high-scoring hits (49).
  - 5) Select best 5 for synthesis "taking into account their individual in silico ranks and building block availability"; reject any if found in chemical or patent databases to ensure novelty.
  - 6) Test against targets in hybrid reporter gene assays (with comparative controls).

**(See the original article for the Results and Discussion)**

I think, the results are inspiring and understandable for a non-AI medicinal chemist. Still, knowing a bit more about the basics of machine learning and neural network architectures is important for life science professionals. So here is a very informative educational piece about all these things to gain the high-concept understanding.