ML for Scent

Alex Wiltschko, Benjamin Sanchez-Lengeling, Brian Lee, Carey Radebaugh, Emily Reif, Jennifer Wei

Hi!

I'm **Alex Wiltschko**, a scientist at Google Research.

I lead a research group within **Google Brain**

that focuses on **machine learning for olfaction**.

Google Research

3500 Researchers & Engineers
18 offices, 11 countries

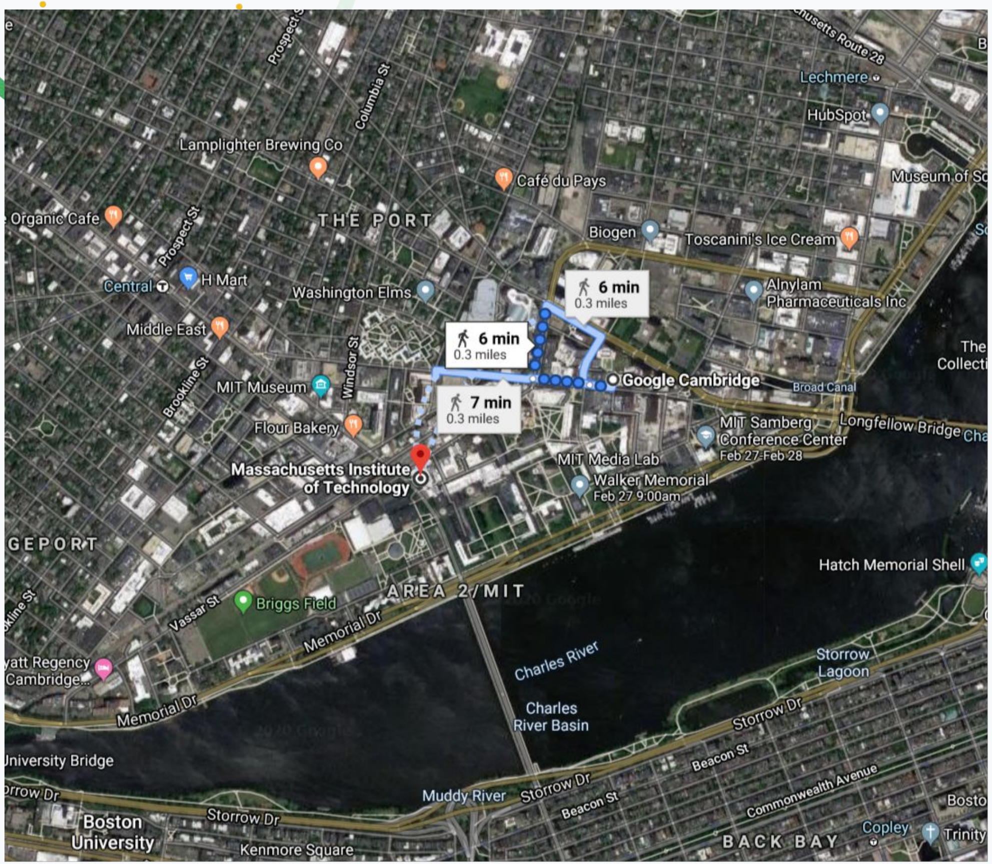
Make machines intelligent. Improve people’s lives.

Google AI

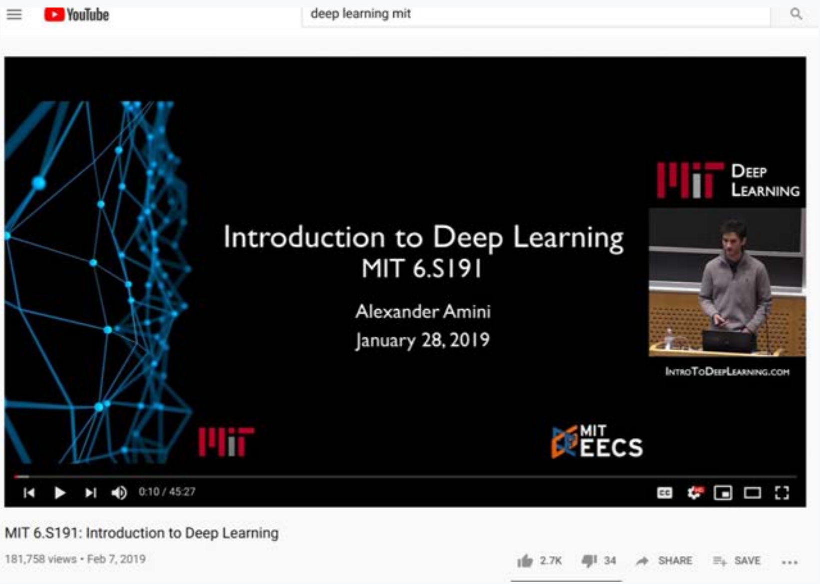
Our Approach

- Foundational research
- Building tools to enable research & democratize AI/ML
- AI-enabling Google products

Google AI



Lechmere
HubSpot
Lamplighter Brewing Co
Café du Pays
Museum of Sc
Organic Cafe
THE PORT
Biogen
Toscanini's Ice Cream
Central
H Mart
Washington Elms
Alnylam Pharmaceuticals Inc
6 min
0.3 miles
Middle East
6 min
0.3 miles
The Collecti
MIT Museum
Google Cambridge
Broad Canal
7 min
0.3 miles
Flour Bakery
MIT Samberg Conference Center
Feb 27-Feb 28
Longfellow Bridge
Massachusetts Institute of Technology
MIT Media Lab
Walker Memorial
Feb 27 9:00am
GEPORT
Hatch Memorial Shell
Briggs Field
AREA 2/MIT
Vassar St
Memorial Dr
Charles River
Storrow Lagoon
Charles River Basin
Memorial Dr
Storrow Dr
Beacon St
University Bridge
Storrow Dr
Muddy River
Storrow Dr
Beacon St
Commonwealth Avenue
Boston University
Kenmore Square
BACK BAY
Copley
Trinity



YouTube
deep learning mit
Introduction to Deep Learning
MIT 6.S191
Alexander Amini
January 28, 2019
MIT Deep Learning
IntroToDeepLearning.com
MIT EECS
0:10 / 45:27
MIT 6.S191: Introduction to Deep Learning
181,758 views • Feb 7, 2019
2.7K
34
SHARE
SAVE

What’s our goal?

Do for olfaction what machine learning has already done for vision and hearing.

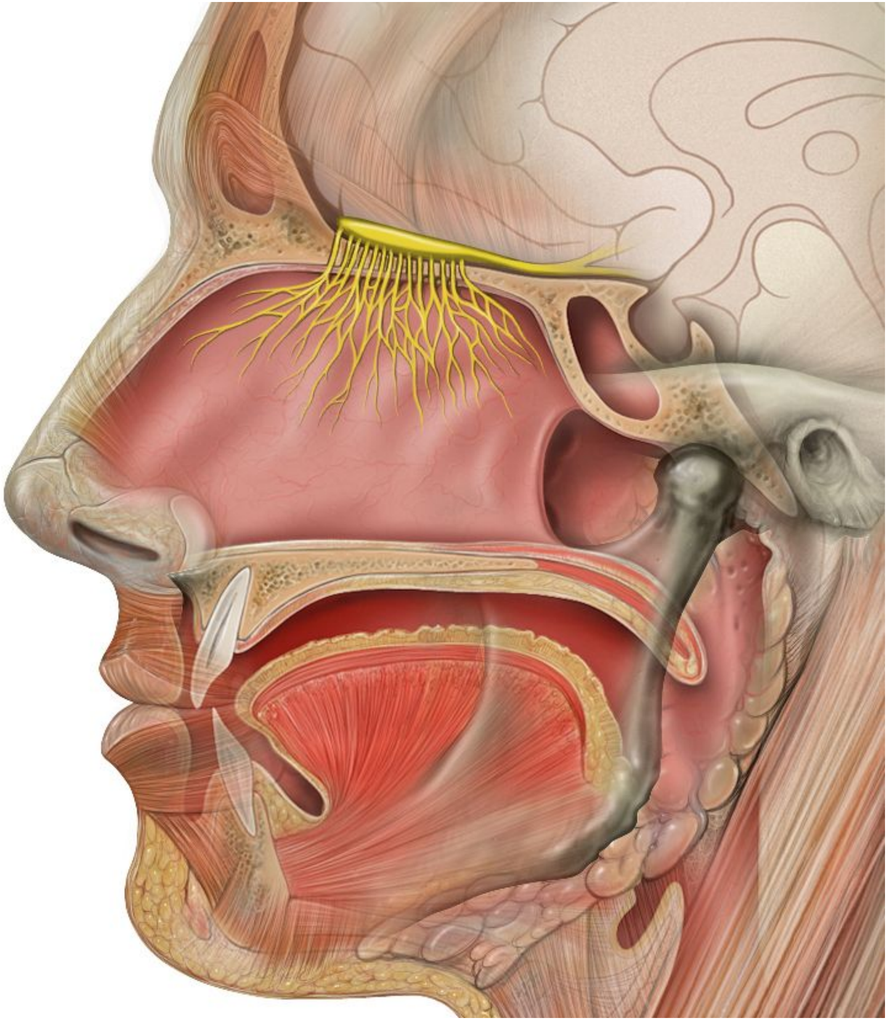
To **digitize the sense of smell**, and make the world’s smells and flavors searchable. Every flower patch, every natural gas leak, every item on every menu in every restaurant.

We’re starting at the very beginning, with the simplest problem…
but first, some olfaction facts!

Most airflow is not smelled. Passes right on through the lower turbinates to your lungs.

The OSNs are one of two parts of your brain that are exposed to the world (the other is the pituitary gland, and that’s in blood, so only half-counts).

Taste lives on your tongue. Flavor is both taste and retronasal olfaction, from a “chimney effect”.



GPCR: G-protein coupled receptor
OR: GPCR Olfactory Receptor
OSN: Olfactory sensory neuron

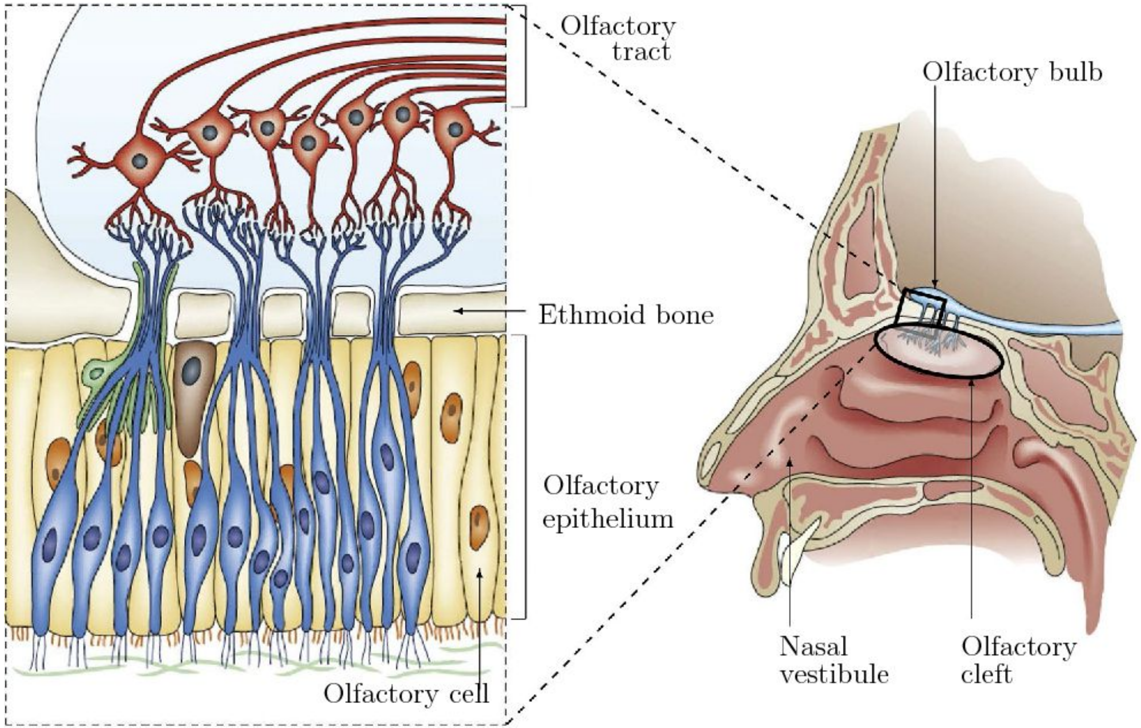
~400 ORs expressed in humans (as opposed to 3 types of cones)
~1000 in mice. ~2000 in elephants!

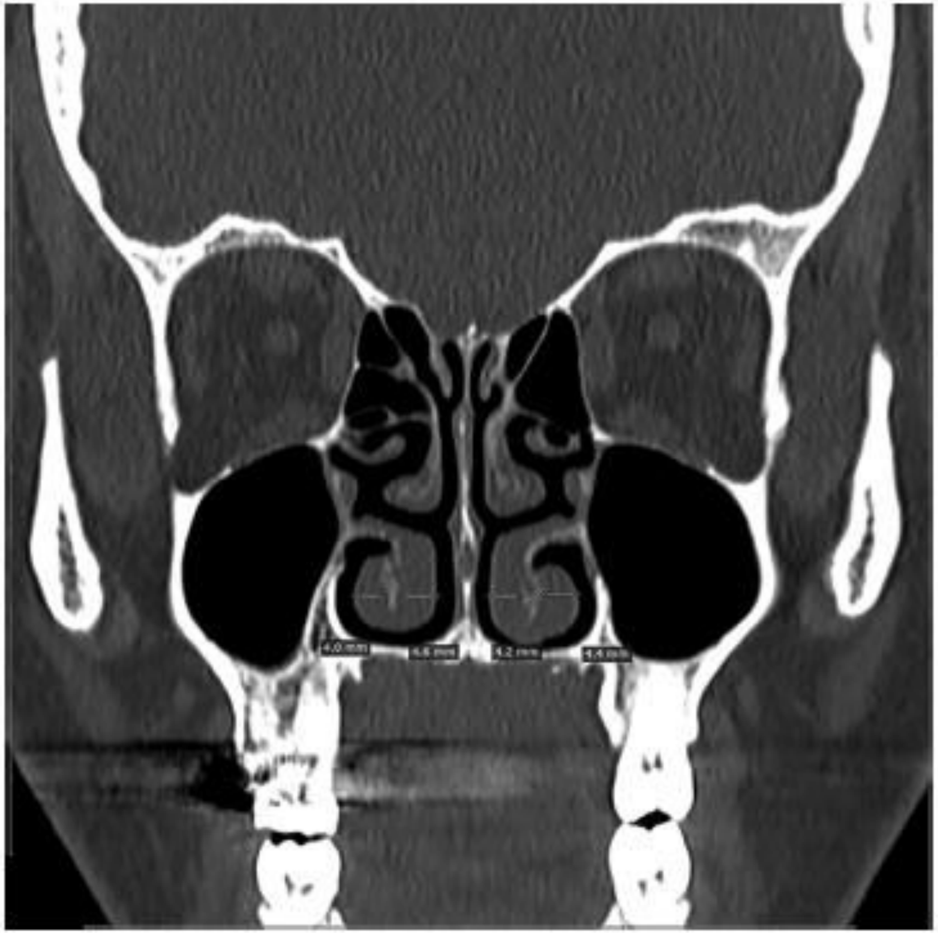
One OR per OSN.

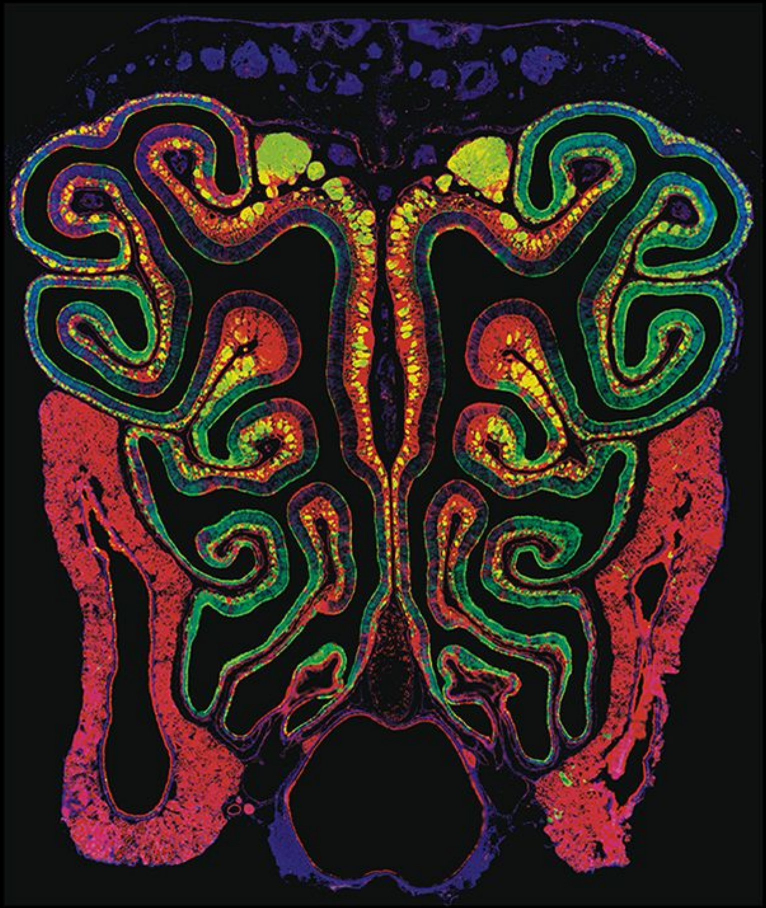
ORs comprise 2% of your genome, but many are pseudogenes.

OR structure is unknown, they are uncrystallized. Further, only ~40 expressed in cell lines.

Their ligand responses are broadly tuned, but many ORs (22/400) are still orphans, with no known ligand.




4.0 mm
4.6 mm
4.2 mm
4.4 mm



People do smell different things!

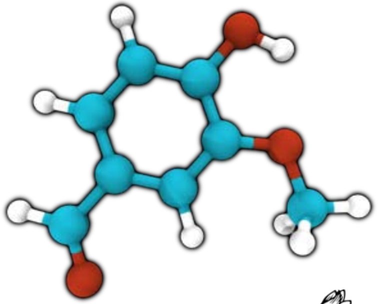
SNPs in single ORs result in sensory dimorphisms. The most famous ones are:

- OR7D4 T113M: normally funky beta-androstenone (boar taint) is rendered pleasant.
- OR5A1 N183D: nearly completely Mendelian. Carriers of the mutation can detect beta-ionine at two orders of magnitude lower concentration
- Olfactory sensory dimorphisms are likely common — humans differ functionally at 30% of OR alleles.
- ~4.5% of the world is colorblind (CBA)
- 13% in the US has selective hearing loss (NIDCD)
- All this to argue — smell is not defacto finicky or illogical.

Right now, we're starting with the ***simplest problem***

Mainland et al 2015

“Smells sweet, with a hint of vanilla, some notes of creamy and back note of chocolate.”



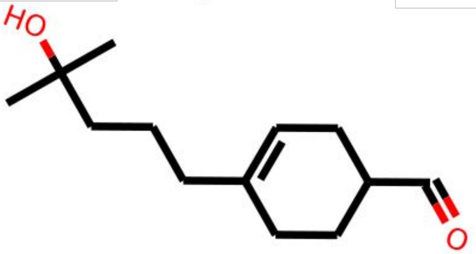
Predict

citrus creamy
sweet baked spicy
odorless vanilla
clean musky beefy
chocolate fruity

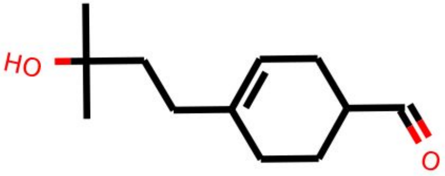
Odor descriptors

And why is this hard?

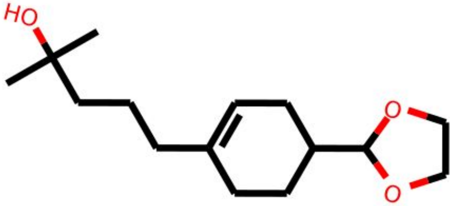
Lyral



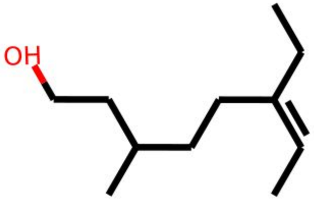
muguet, fresh, floral, sweet



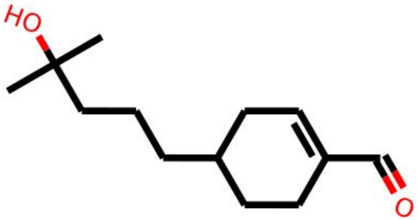
muguet floral



muguet, fresh, floral, sweet

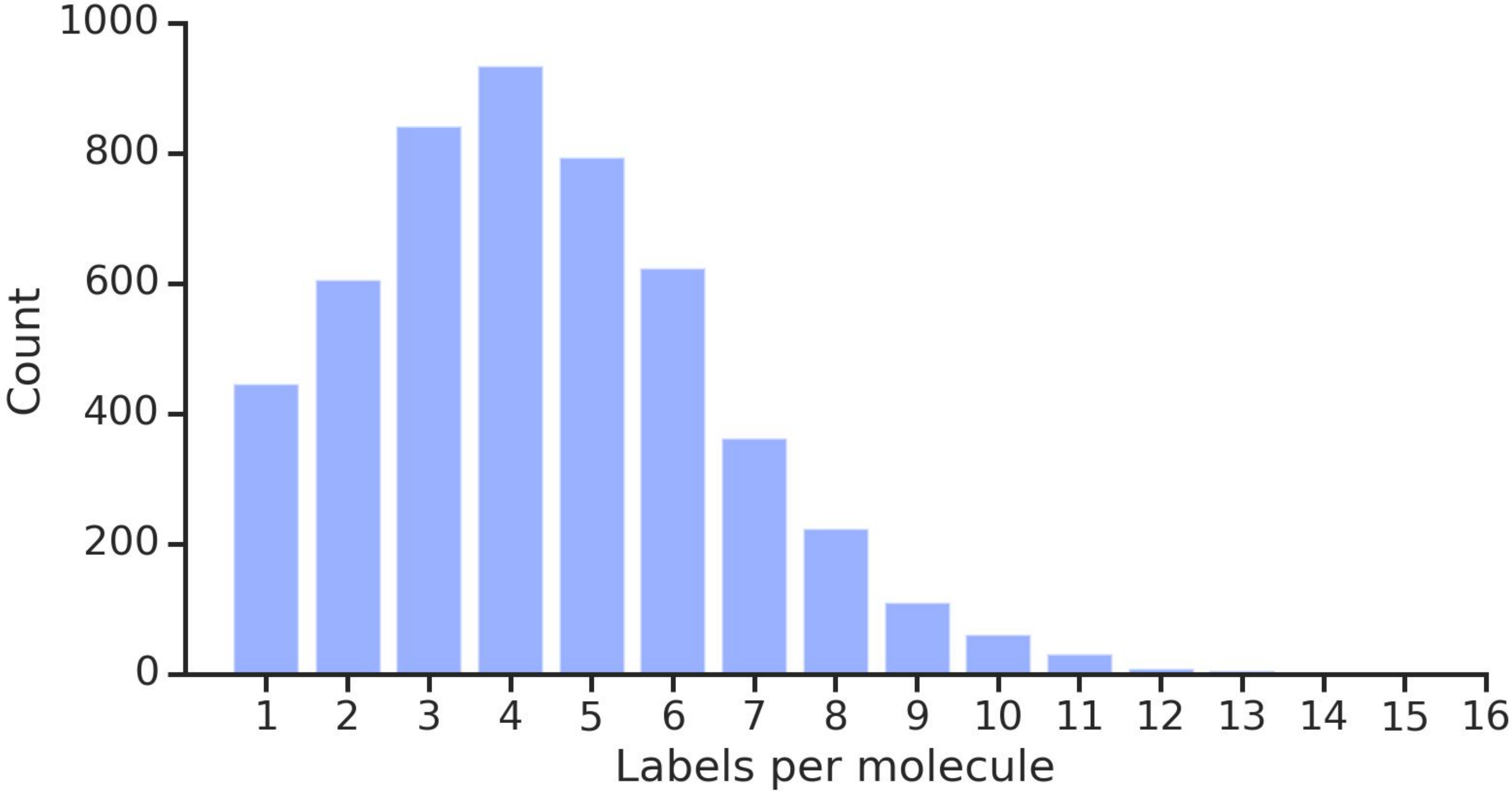


muguet, fresh, floral, rosy

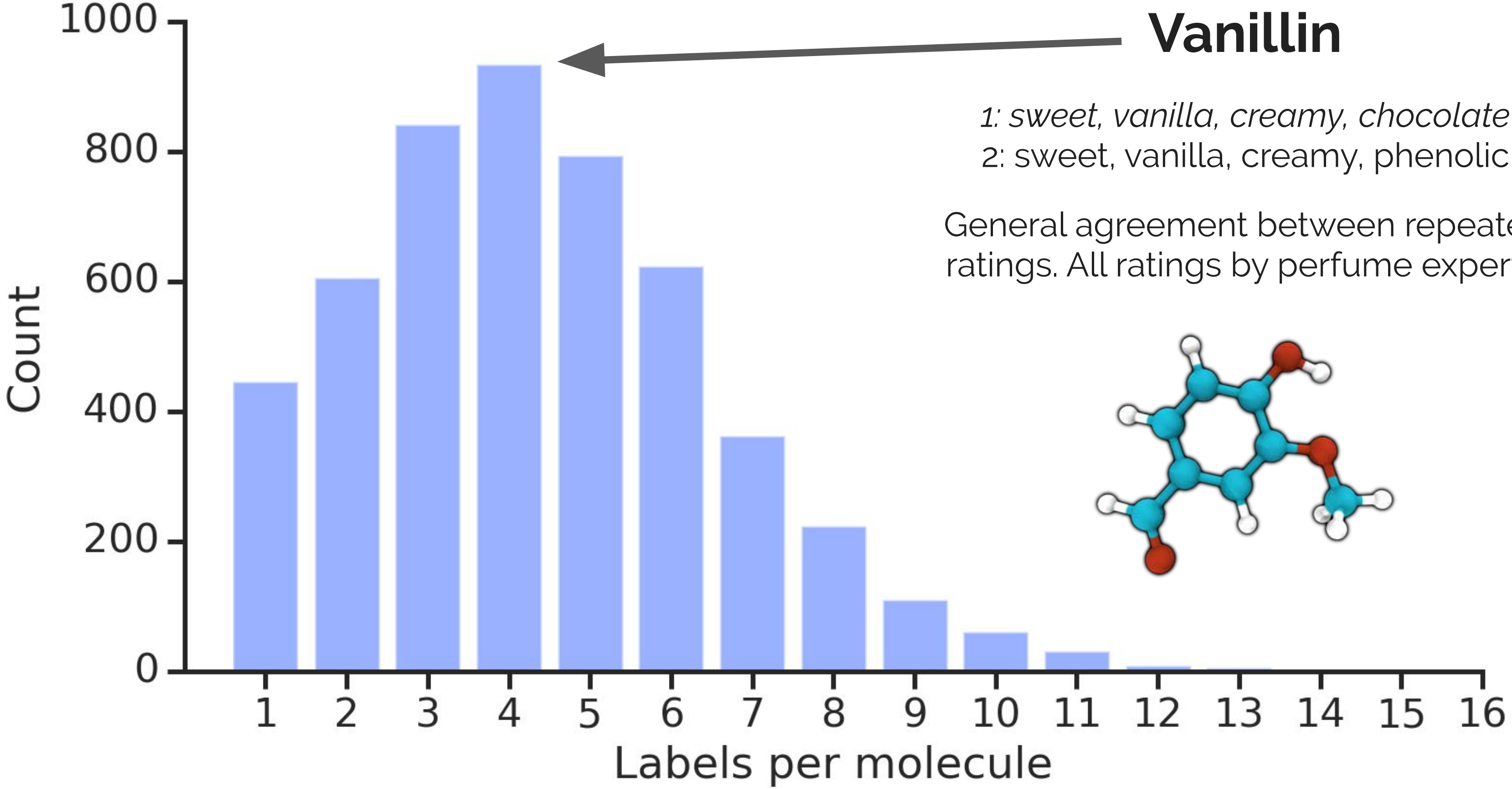


odorless

We built a benchmark from perfumery raw materials



We built a benchmark from perfumery raw materials

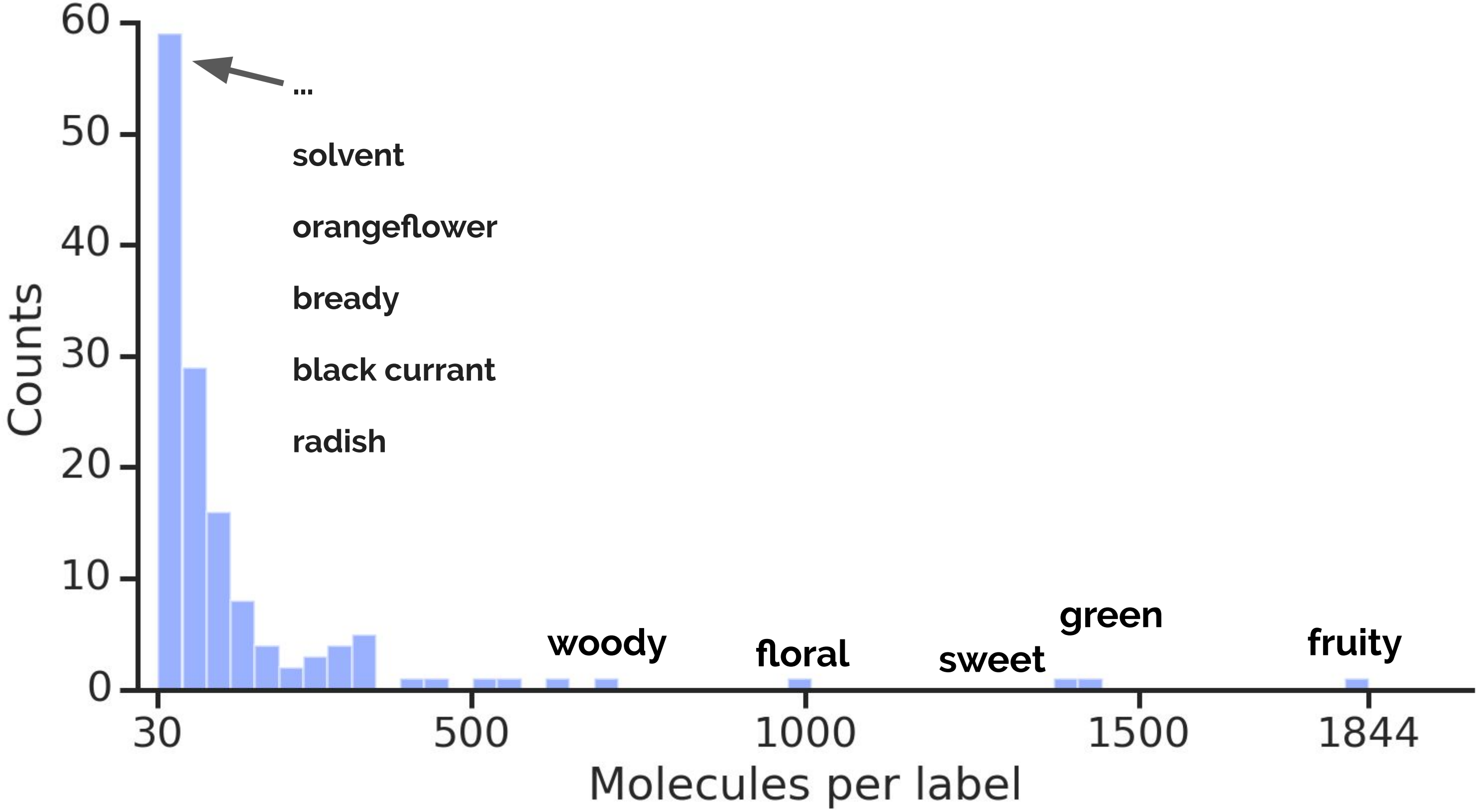


Vanillin

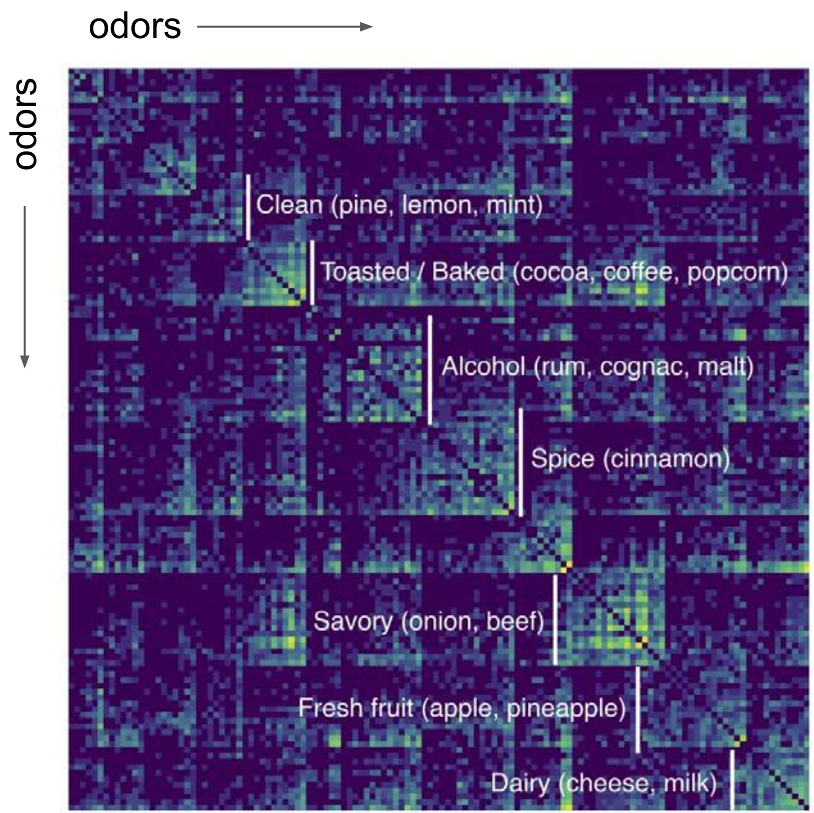
1: sweet, vanilla, creamy, chocolate
2: sweet, vanilla, creamy, phenolic

General agreement between repeated ratings. All ratings by perfume experts.

We built a benchmark from perfumery raw materials



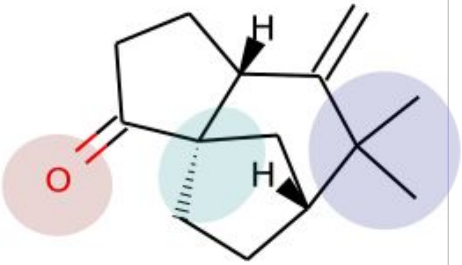
We built a benchmark from perfumery raw materials



Historical SOR approaches

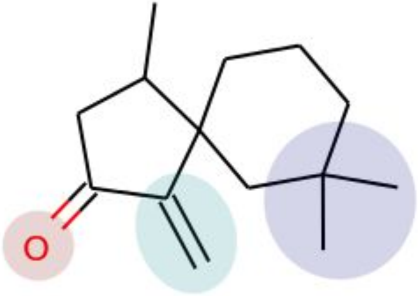
Pen & Paper

Ohloff's rule
Bajgrowicz and Broger's ambergris osmophore model
Buchbauer's santalols
Boelens' synthetic muguet

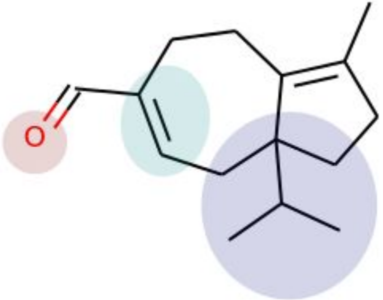


(-)-khusimone

Kraft's vetiver rule



4,7,7-Trimethyl-1-methylidene spiro[4.5]decan-2-one

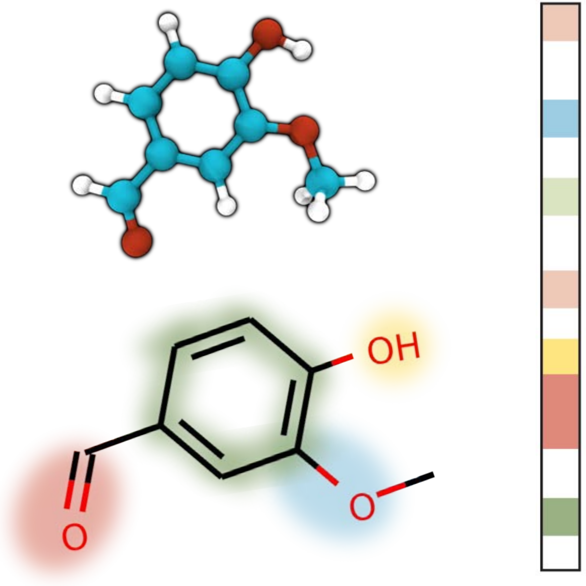


1,7-cyclogermacra-1 (10),4-dien-15-al

Fig 3.22 *Scent and Chemistry* (Ohloff, Pickenhagen, Kraft)

Rule-based principles for predicting odor. There are as many exceptions as there are rules.

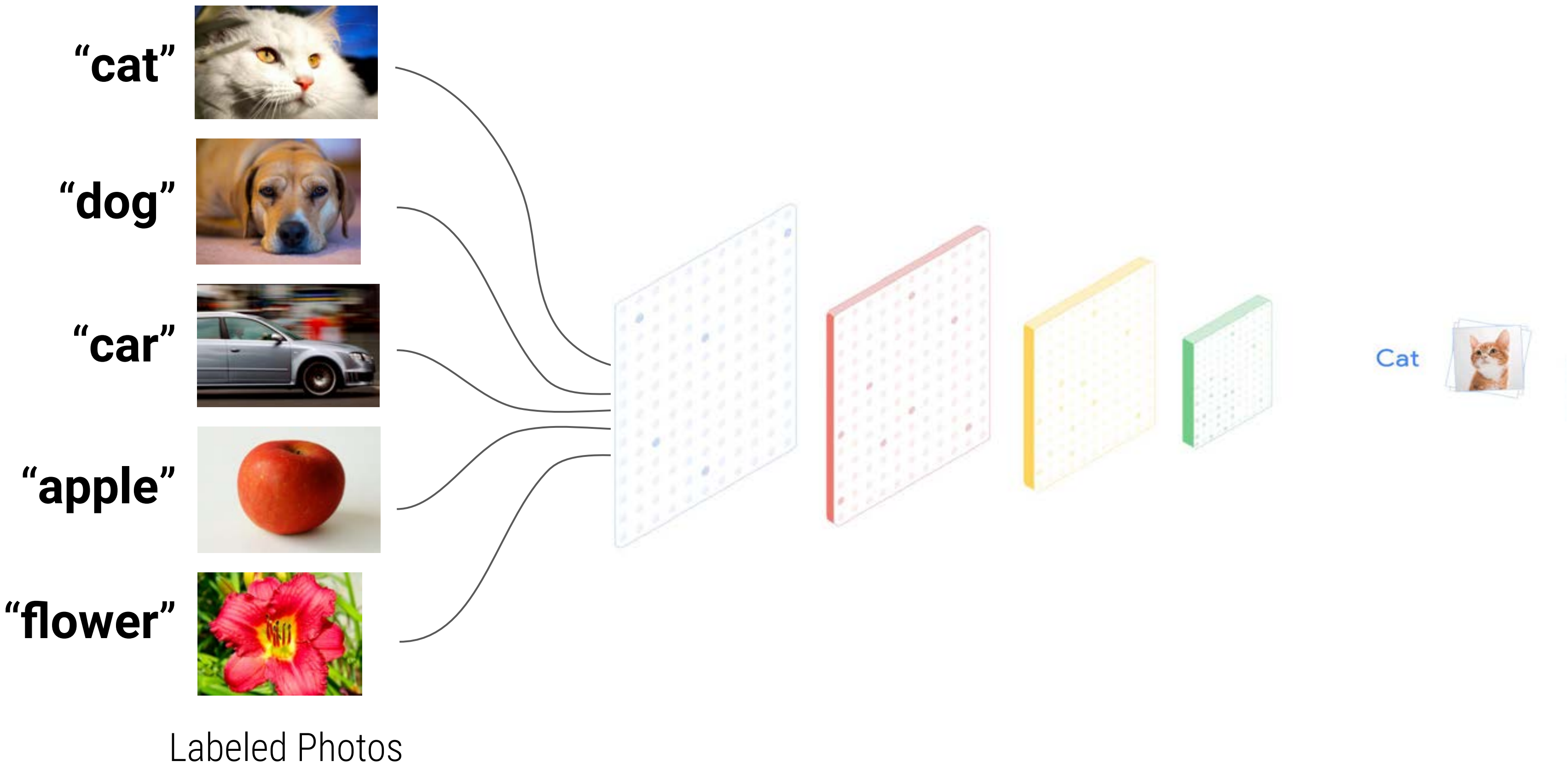
Traditional Computational Approaches

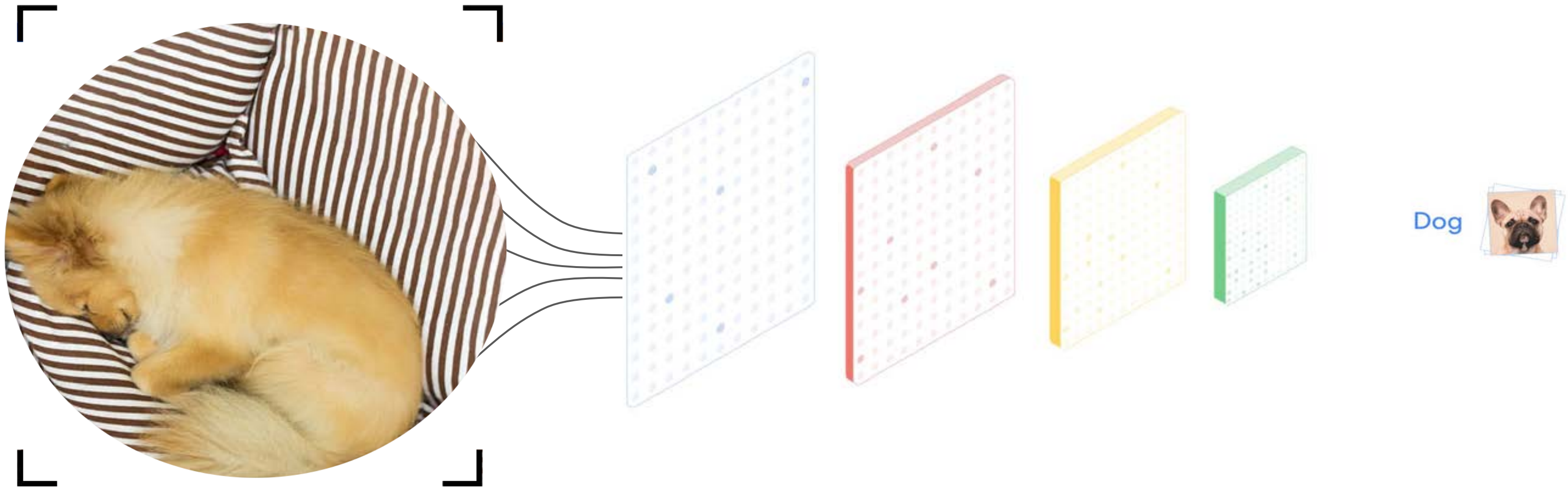


“bag of sub-graphs” representation
AKA molecular fingerprints

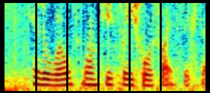
Predict →

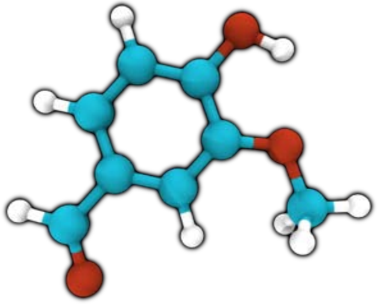
- Toxicity
- Solubility
- Photovoltaic efficiency (solar cell)
- Chemical potential (batteries)
- ...


"cat"
"dog"
"car"
"apple"
"flower"
Labeled Photos
Cat



Unlabeled Photo

Input		Output
	PIXELS	“lion”
	AUDIO	“How cold is it outside?”
“Hello, how are you?”	TEXT	“你好，你好吗？”
	PIXELS	“A blue and yellow train travelling down the tracks”



Graphs as input to neural networks: not just images, sounds or words

Convolutional Networks on Graphs for Learning Molecular Fingerprints

David Duvenaud†, Dougal Maclaurin†, Jorge Aguilera-Iparraguirre
Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, Ryan P. Adams
Harvard University

Abstract

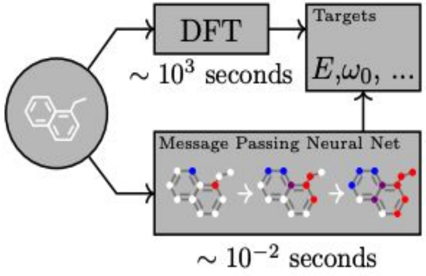
We introduce a convolutional neural network that operates directly on graphs. These networks allow end-to-end learning of prediction pipelines whose inputs are graphs of arbitrary size and shape. The architecture we present generalizes standard molecular feature extraction methods based on circular fingerprints. We show that these data-driven features are more interpretable, and have better predictive performance on a variety of tasks.

Neural Message Passing for Quantum Chemistry

Justin Gilmer [1] Samuel S. Schoenholz [1] Patrick F. Riley [2] Oriol Vinyals [3] George E. Dahl [1]

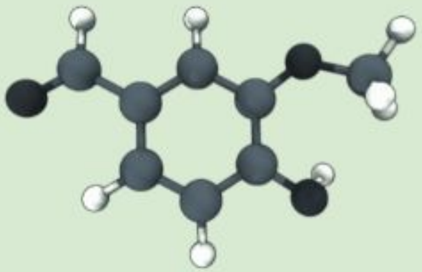
Abstract

Supervised learning on molecules has incredible potential to be useful in chemistry, drug discovery, and materials science. Luckily, several promising and closely related neural network models invariant to molecular symmetries have already been described in the literature. These models learn a message passing algorithm and aggregation procedure to compute a function of their entire input graph. At this point, the next step is to find a particularly effective variant of

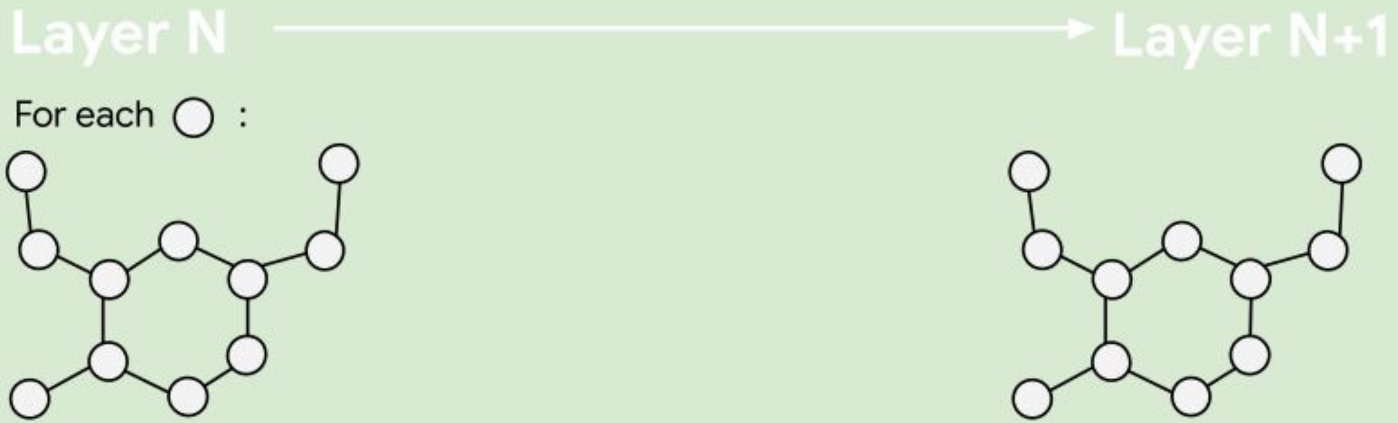


Inside a GNN
Converting a molecule to a graph

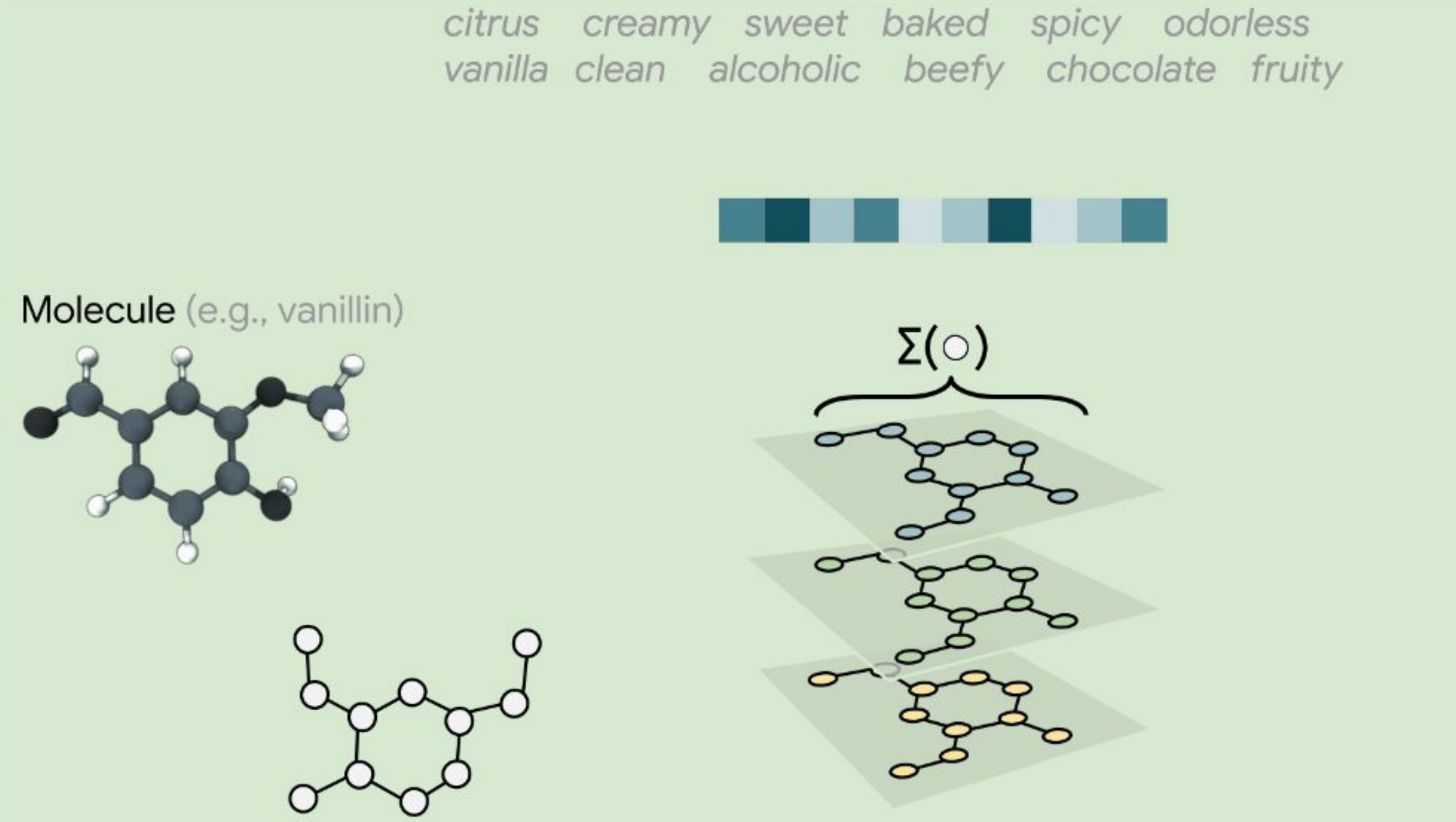
Molecule (e.g., vanillin)



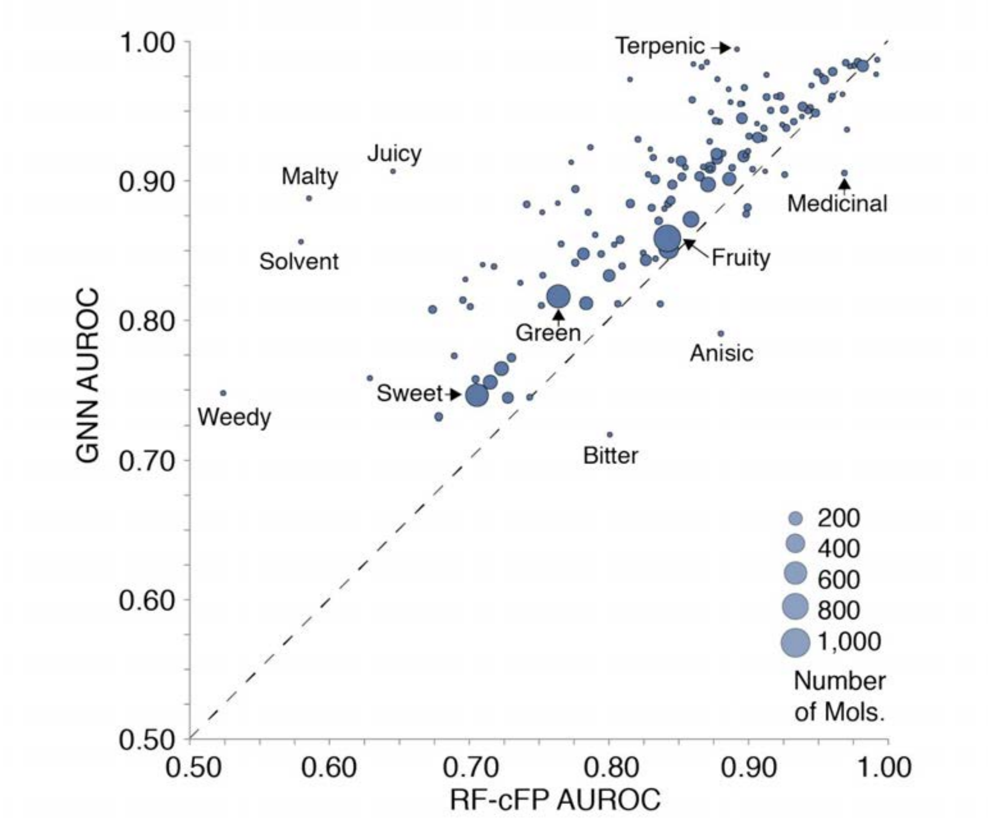
Inside a GNN
Propagating information & transforming a graph



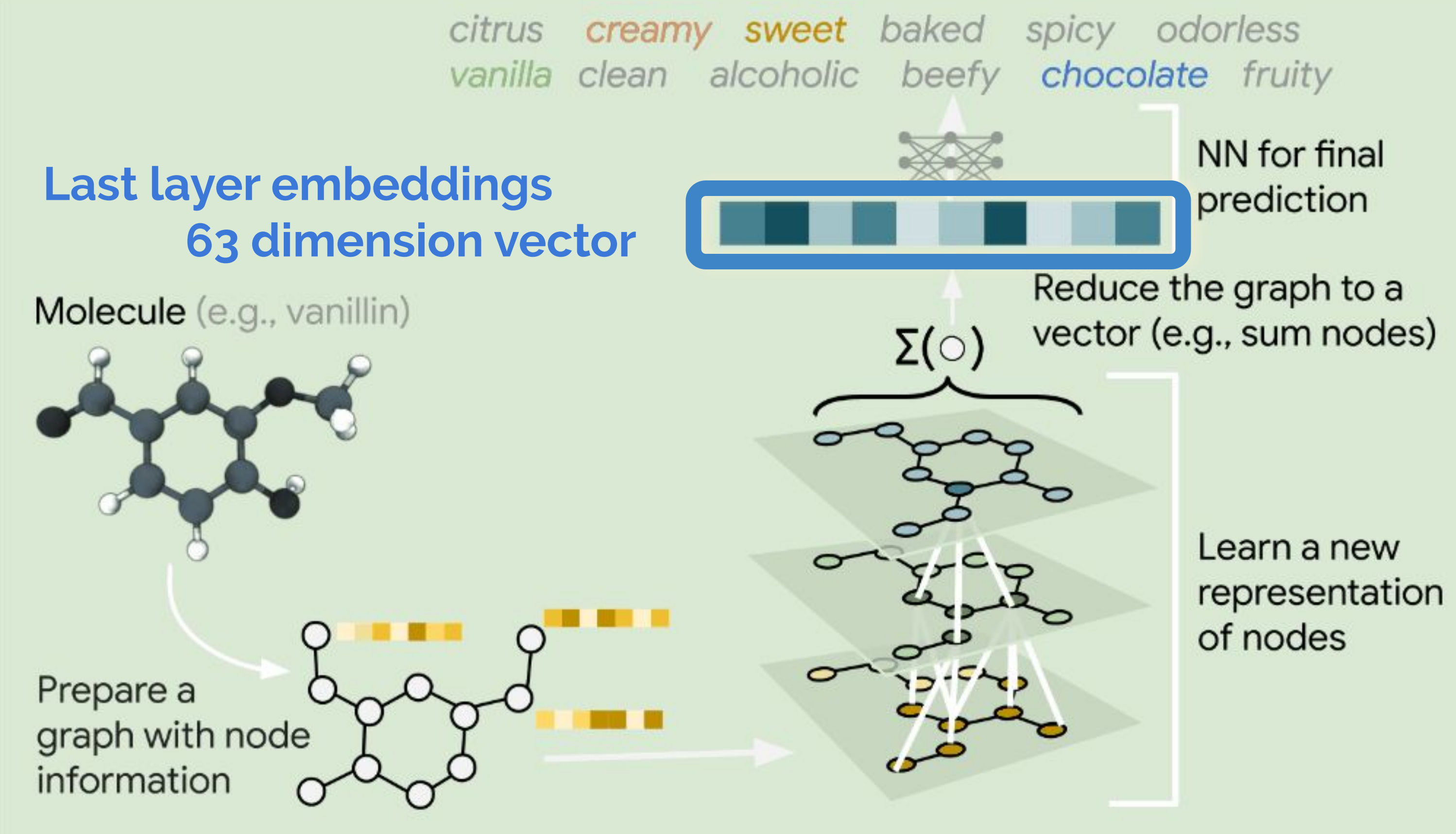
A GNN to predict odor descriptors



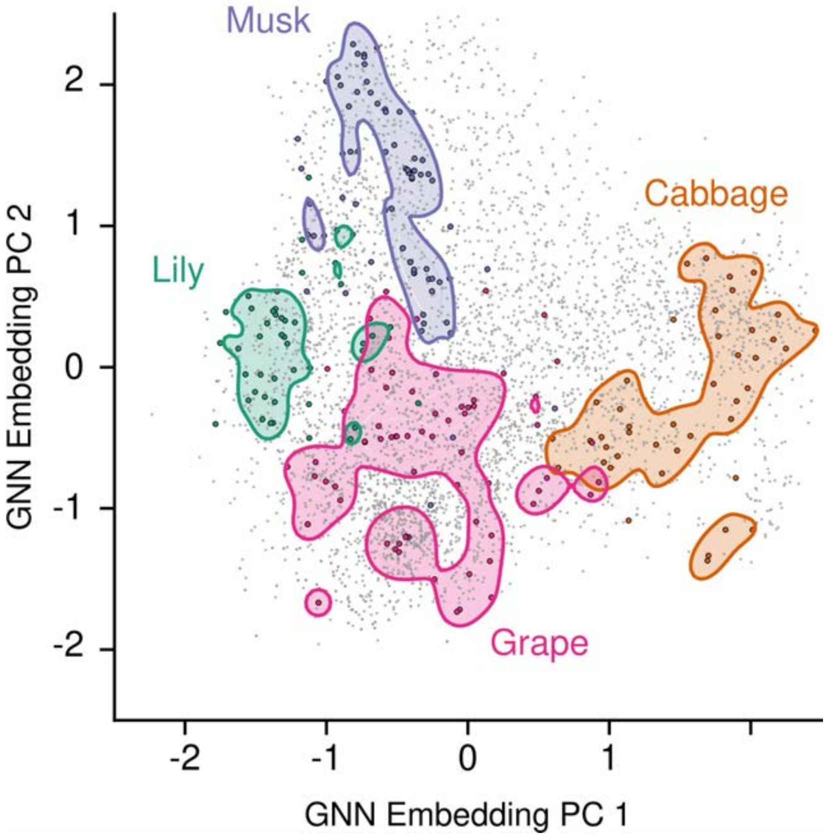
And how well can we predict?



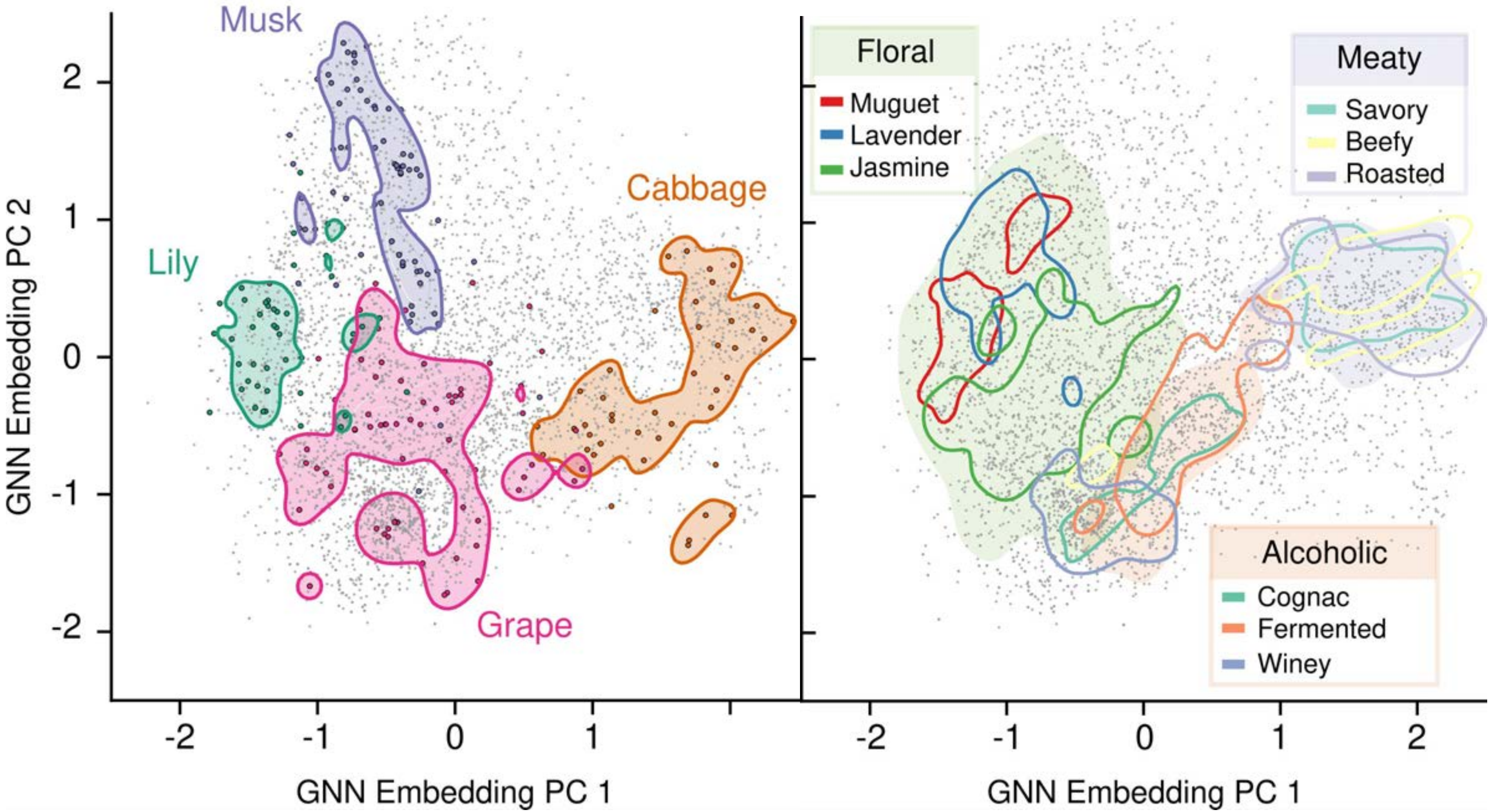
A representation optimized for odor



Exploring the geometric space of odor



Exploring the geometric space of odor



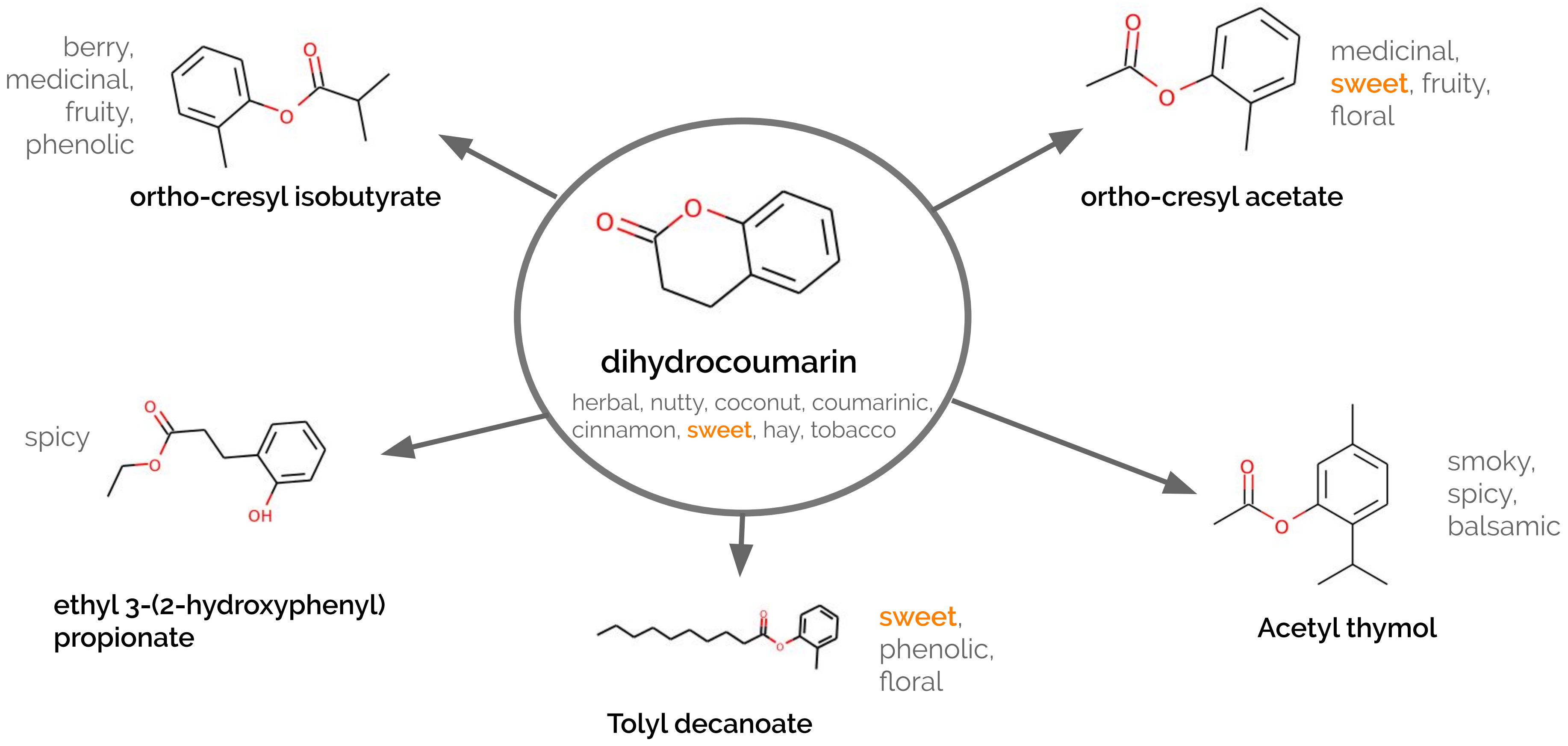
What do nearby molecules look like?

Inspired by word embeddings. Are there “molecular synonyms”?

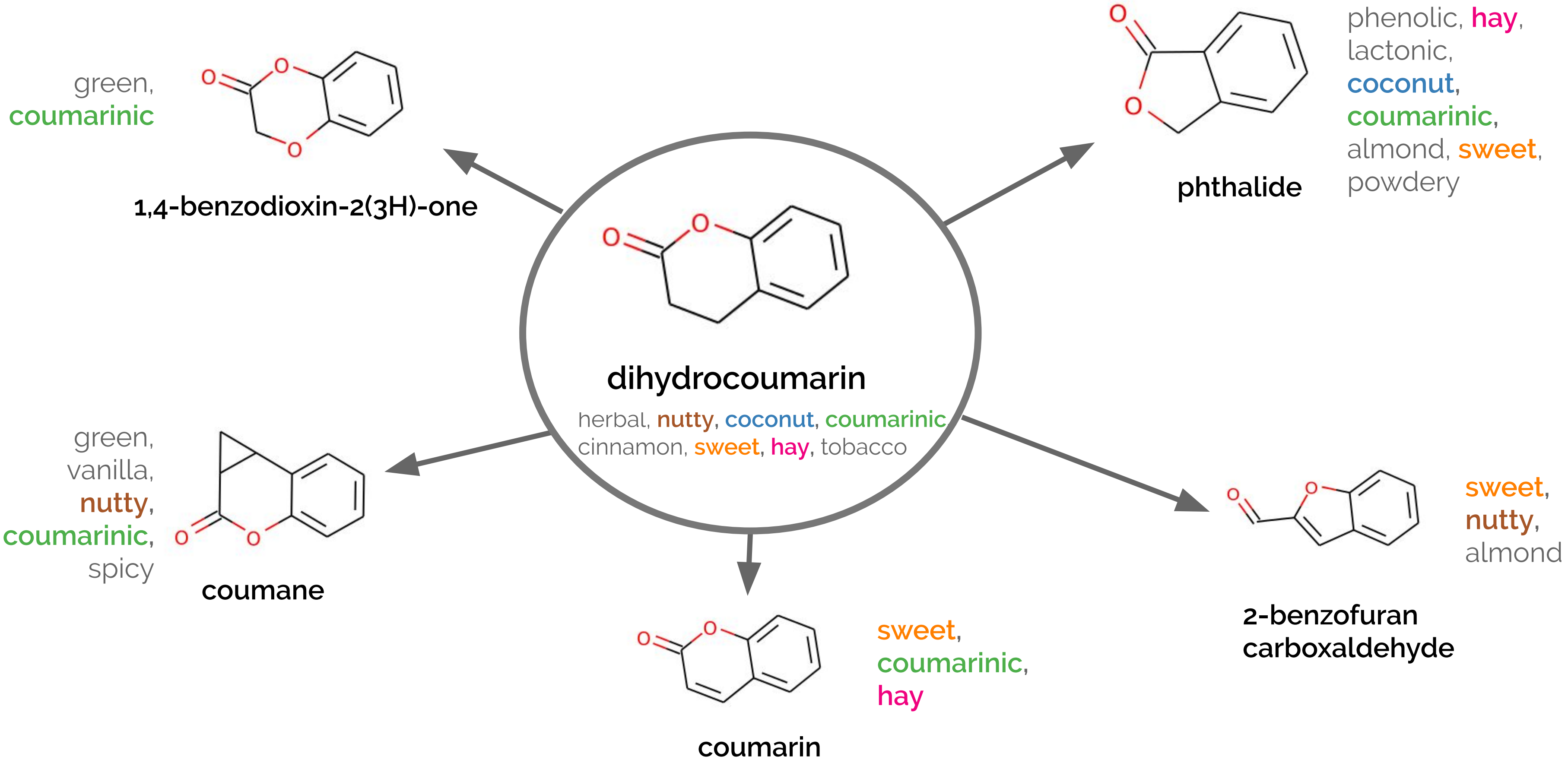
First, what do “nearest neighbors” look like if you use just structure, and ignore our neural network?

Then, what do nearest neighbors look like to our GCN?

Molecular neighbors: using structure



Molecular neighbors: using GCN features



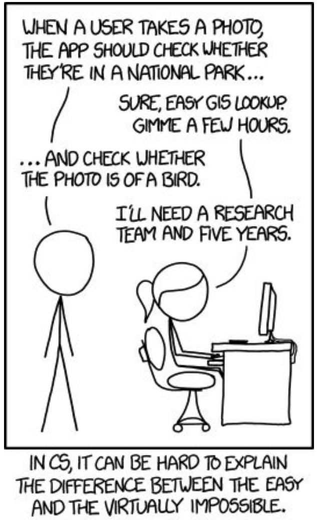
Do these representations generalize?

Using a learned model to make predictions on a new task is ‘transfer learning’

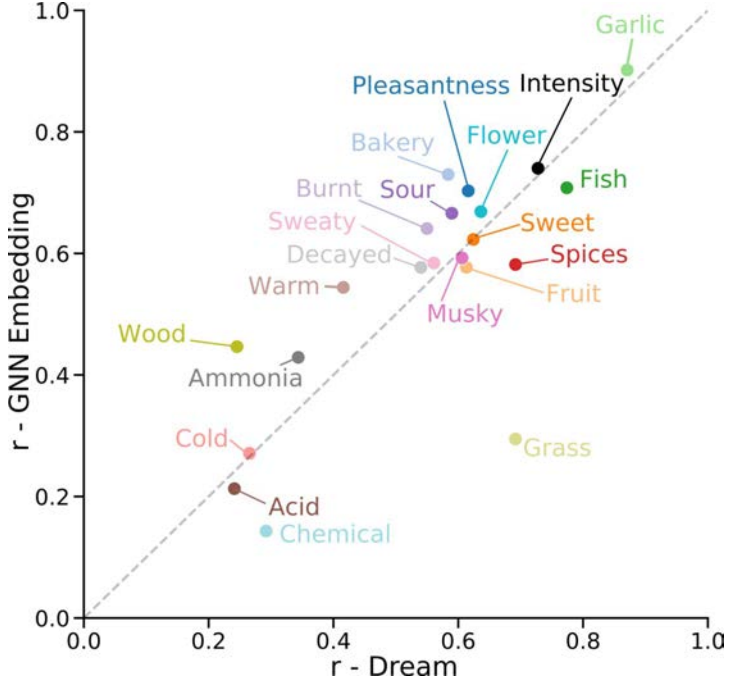
You might hear ‘fine-tuning’ referred to as a strategy for ‘transfer learning’.

Transfer learning in chemistry, today, rarely works. Do our embeddings transfer learn to other tasks?

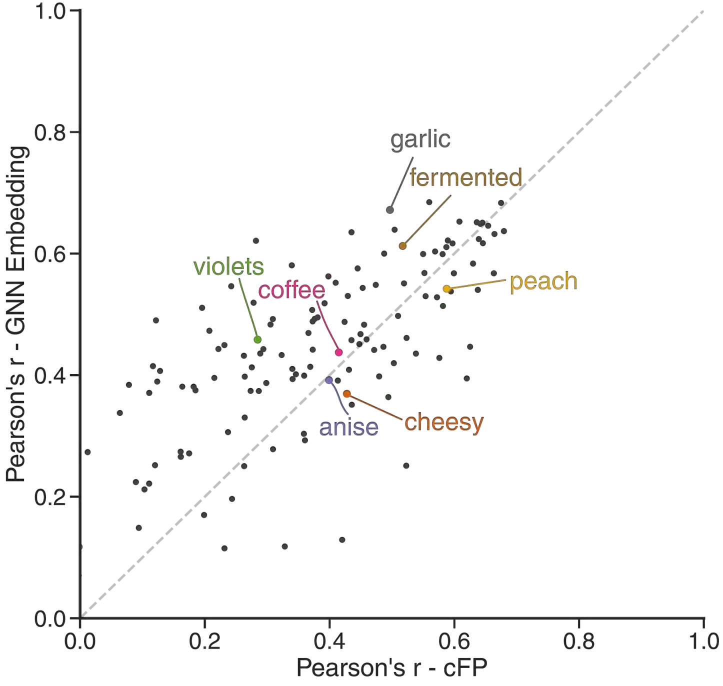
Do these representations generalize?



DREAM Olfactory Challenge



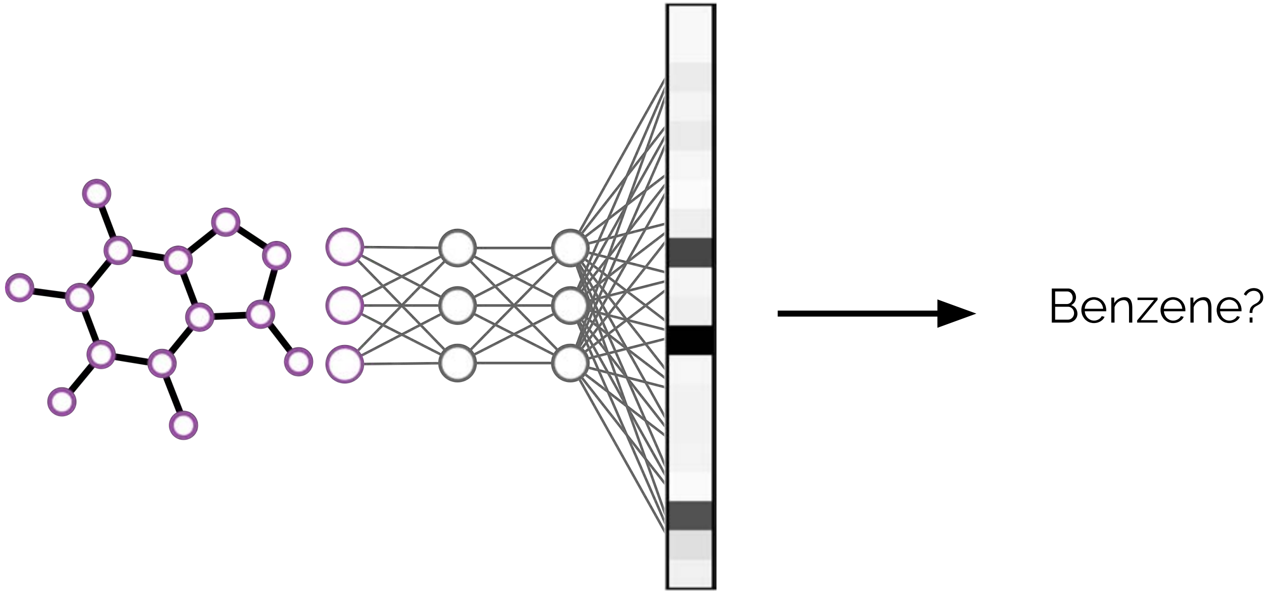
Dravnieks



Transfer-learned to achieve state-of-the-art on the two major olfactory benchmark tasks

But *why* is the neural network making these predictions?

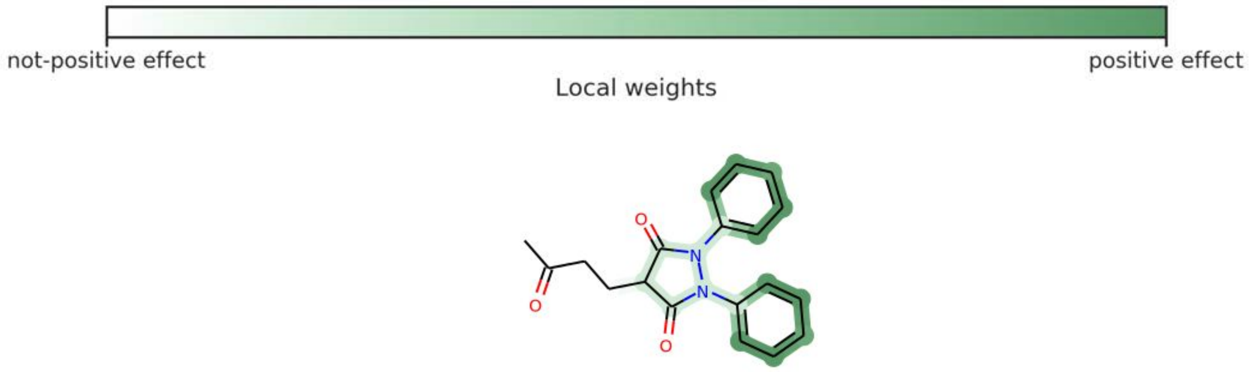
Toy test example: classify whether a molecule has benzene. Which atoms contribute to predictions?



This is just one task of potentially hundreds, of varying complexity.

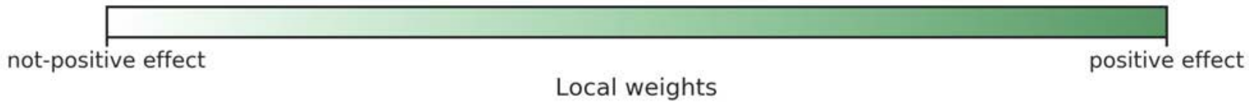
But *why* is the neural network making these predictions?

Toy test example: classify whether a molecule has benzene. Which atoms contribute to predictions?

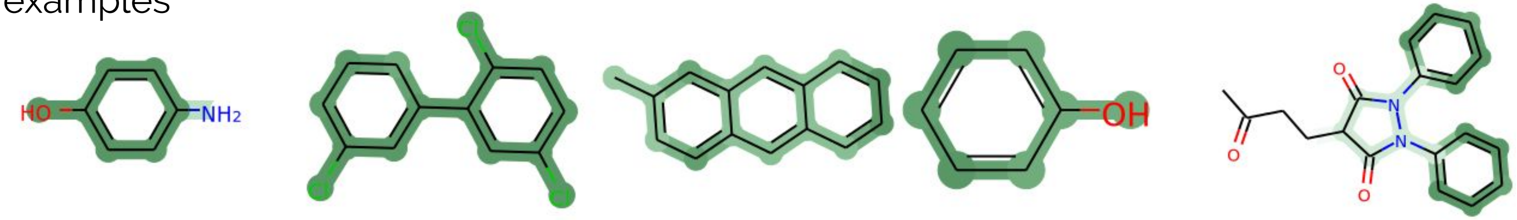


But *why* is the neural network making these predictions?

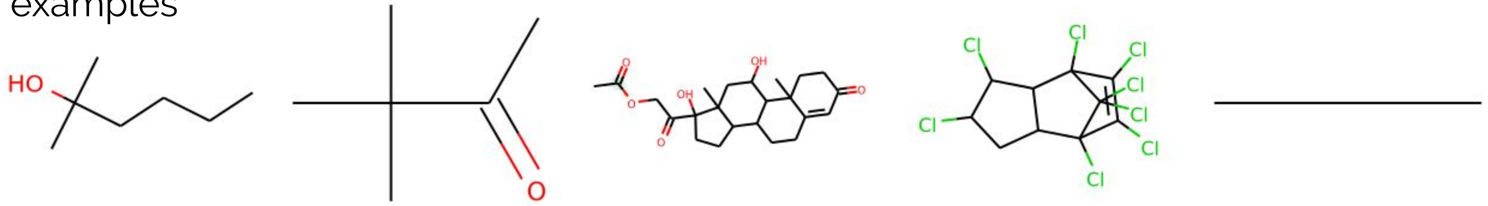
Toy test example: classify whether a molecule has benzene. Which atoms contribute to predictions?



Positive examples

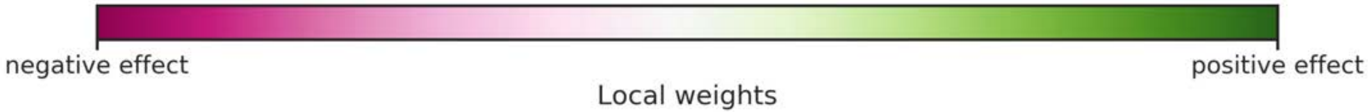


Negative examples



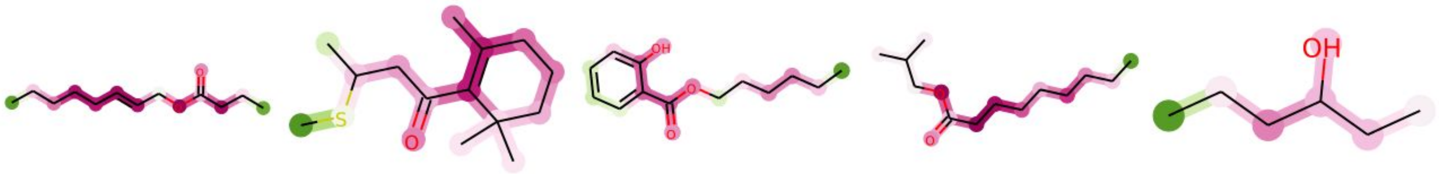
But *why* is the neural network making these predictions?

Odor percept — “garlic”



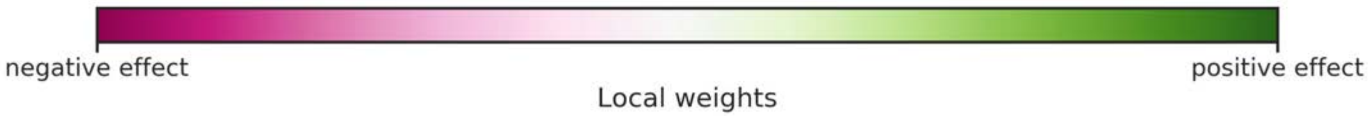
Positive examples

Negative examples



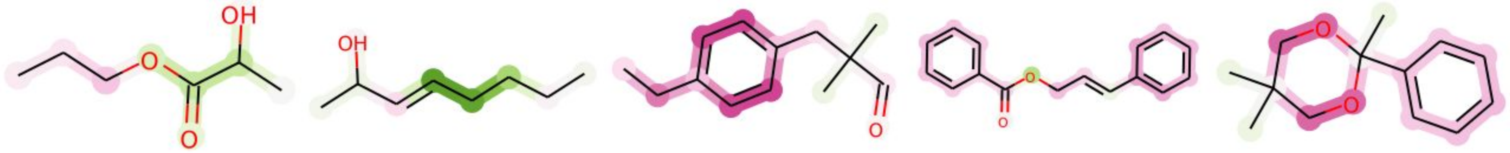
But *why* is the neural network making these predictions?

Odor percept — “fatty”



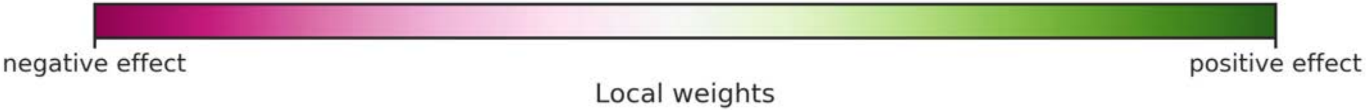
Positive examples

Negative examples

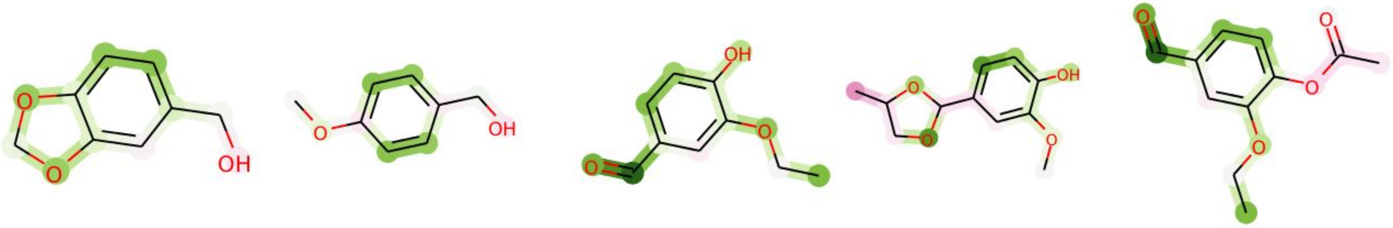


But *why* is the neural network making these predictions?

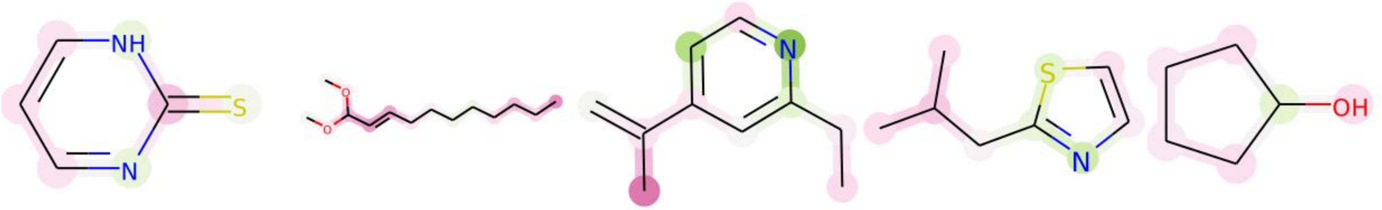
Odor percept — “vanilla”



Positive examples

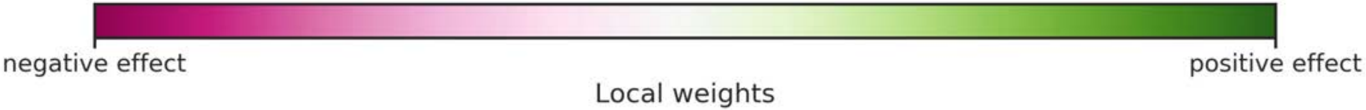


Negative examples

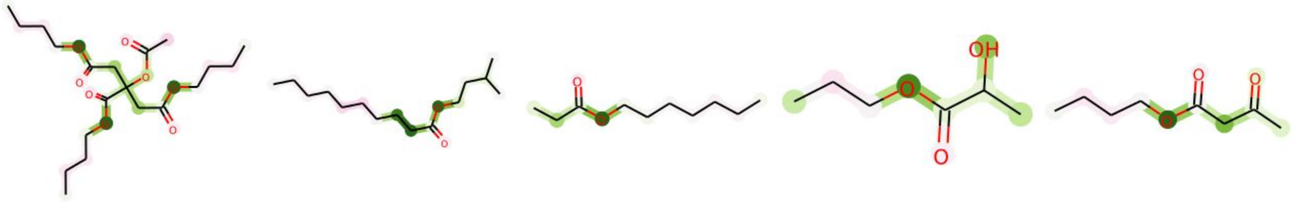


But *why* is the neural network making these predictions?

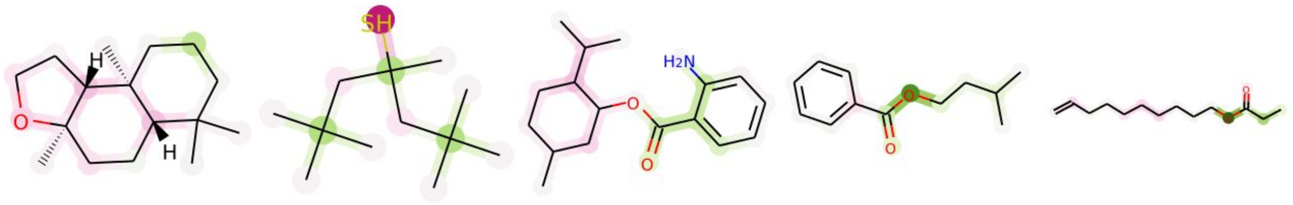
Odor percept — “winey”

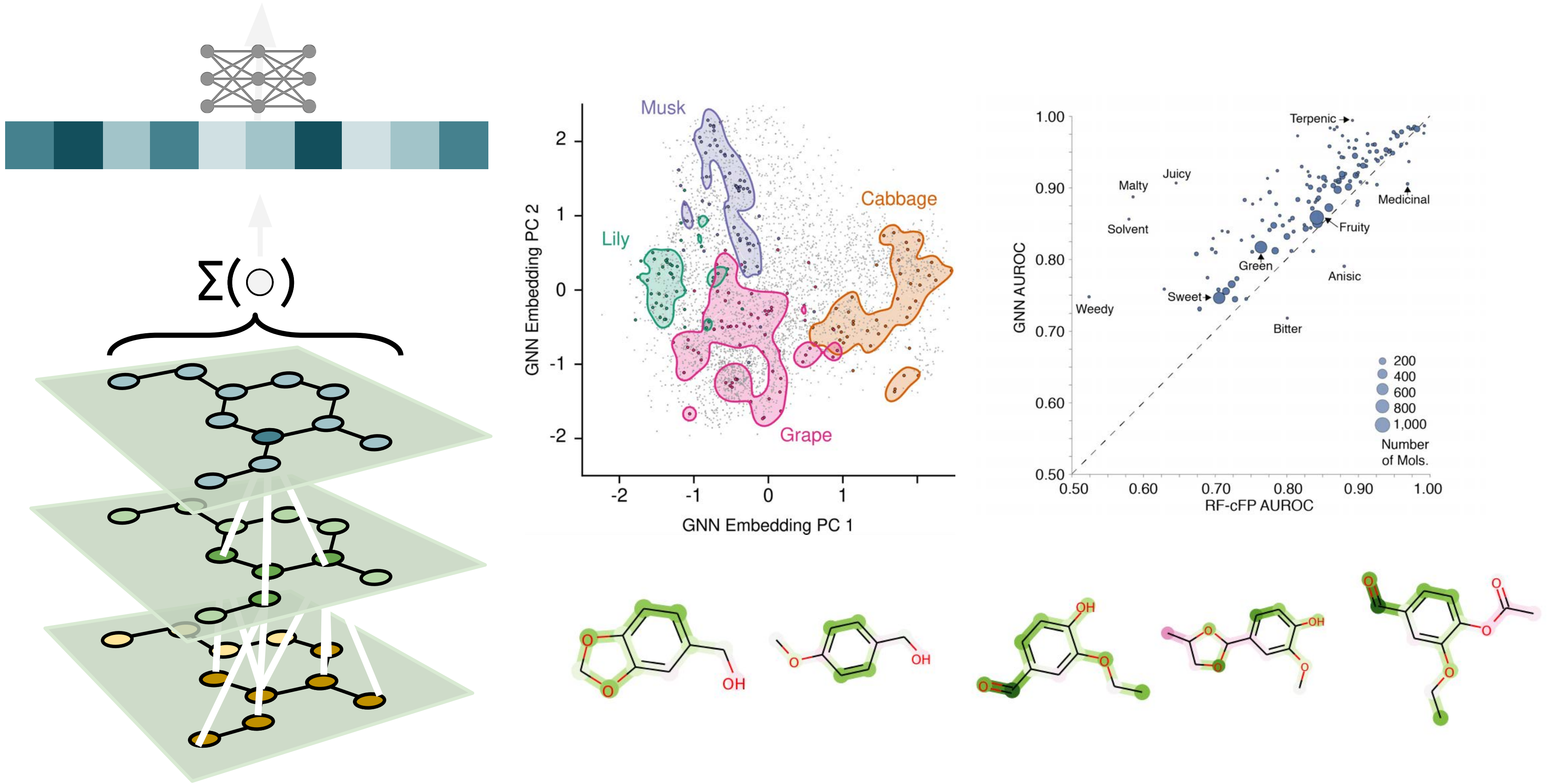


Positive examples



Negative examples





Σ(○)
Musk
Lily
Cabbage
Grape
GNN Embedding PC 2
GNN Embedding PC 1
2
1
0
-1
-2
-2
-1
0
1
1.00
0.90
0.80
0.70
0.60
0.50
GNN AUROC
0.50
0.60
0.70
0.80
0.90
1.00
RF-cFP AUROC
Terpenic
Juicy
Malty
Medicinal
Solvent
Fruity
Green
Anisic
Sweet
Weedy
Bitter
200
400
600
800
1,000
Number of Mols.
O
O
OH
O
OH
OH
O
O
O
O
OH
O
O
O
O
O
O

Future Directions

Collecting interest & those interested in collaborating.

- **Test ML-driven molecular design** for humans in a safe context.

- Build bedrock understanding in single-molecules before working on **odor mixtures**

- Build a **foundational dataset** for the ML on molecules community.

Benjamin Sanchez-Lengeling
Brian Lee
Carey Radebaugh
Emily Reif
Jennifer Wei
Alex Wiltschko