[DAISY THE GREAT](https://www.bandsintown.com/en/a/14112703-daisy-the-great): 00:00 [Built my home on hollow ground](https://www.youtube.com/watch?v=z-1sC1lkmKw).

CRAIG: 00:06 This is [Craig Smith](https://en.wikipedia.org/wiki/Craig_S._Smith) with a podcast about artificial intelligence. This week I enter the ‘smelliverse,’ the dimension of odors, with [Alexei Koulakov](https://www.cshl.edu/research/faculty-staff/alexei-koulakov/), a professor at Cold Spring Harbor Laboratory on the north shore of Long Island. Alexei uses machine learning and other tools to explore the least understood of our senses: the sense of smell. I’ve talked to people on the podcast about [computer vision](https://www.eye-on.ai/podcast-017), [computer hearing](https://www.eye-on.ai/podcast-009) and even the [sense of touch in robots](https://www.eye-on.ai/podcast-014). But I had never heard anyone talk about using machine learning to explore the sense of smell, until I met Alexei. He’s working to understand the mechanism by which we perceive odors and classify millions of volatile molecules by their smell. It’s a fascinating subject and will have you sniffing the air at odd moments, increasing your awareness of this mysterious dimension. I hope you enjoy listening to Alexei as much as I did.

CRAIG: 01:09 So I thought I would get you to sit down and talk about the physics of odor or the biology of, of odor perception and how you've been able to apply machine learning to classify odors and that sort of thing. But to begin, can, can you sort of tell me about yourself, how you got into this, what's your educational background?

ALEXEI: 01:35 Well, I was trained as applied mathematician, engineer and then I became a theoretical physicist when I moved to the States. And then I have, like, throughout my life, I had a very strong interest in biology, right. And uh, in particular in how the brain works, neuroscience. And for some reason I got lured into the, into neuroscience from, from theoretical physics. Although I did actually hold, like, a faculty position in a, in physics. But I started doing biology, but then I decided that it's better for me, like, basically to switch to biology, to neuroscience completely. So I, I went to [Cold Spring Harbor](https://www.cshl.edu/) to do it, you know, but I do something that is called theoretical neuroscience called theoretical, theoretical/computational neuroscience, which means that, I'm trying to understand what are the models, what are the theoretical models that can explain brain function. So that's the overall idea. And I basically apply methods, like, various methods that I learned in computer science, theoretical physics, math, applied math to understanding the brain function. That's what I do on everyday basis. It's, like, the best job in the world.

CRAIG: 02:47 Yeah.

ALEXEI: 02:48 And I think the most exciting part of that job is that you get to learn a lot because it's a very diverse field. And uh, that's the, basically the thing that drives me.

MUSIC: 02:59 Musical Interlude

CRAIG: 03:08 And then out of that came an interest in olfaction.

ALEXEI: 03:13 So olfaction basically is the least understood of our senses. One would say that vision is now as poorly understood, but, like, high level vision is poorly understood. On the lower level, we know how, you know, photo receptors in the eye, how they interact with light. Right. But in olfaction, we don't have even that basic understanding. So we don't really know, like, the mechanisms by which molecules activate olfactory receptors. I mean olfactory receptors are interesting because they're very similar to the ones that, the photoreceptors in the eye. So they're related. In fact, there are much more olfactory sensors than - we have three color vision, right? So we have three types of photoreceptors in the eye. Um, we have about [350 olfactory receptors](https://en.wikipedia.org/wiki/Olfactory_receptor_neuron). So roughly speaking olfactory system is the retina with 350 types of colors. Wow. And maybe four pixels.

ALEXEI: 04:08 And it's not clear. Like, you know, in the retina we have about a million pixels, right? For olfactory system, we have about four pixels. So the spatial resolution is really limited. But most of the resolution in olfactory system goes into this "color," you know, in quotations signs, that basically allows us to recognize different molecules. Yeah. But what is unclear is what type of molecules, what type of features in the molecules olfactory system is sensitive to.

CRAIG: 04:35 Why would evolution develop kind of an imbalance between what you're calling the pixels and the.

ALEXEI: 04:42 Well, the spatial resolution versus the so-called feature, the feature space versus the spatial. I mean this is not really clear. I think, like, overall that, spatially, olfactory stimuli, which are streams of odors in the environment, they are not so precise, they're very dispersed. But again, I'm speculating here, like, there, there are a lot of interesting interactions between, anecdotal kind of interactions, between the visual system and olfactory system.

ALEXEI: 05:10 Because, like, for example, if you take a mouse it would have about a thousand of olfactory receptors and humans about have about 350. So we lost a lot of olfactory receptors in the course of evolution. And so there is a suggestion which was backed up by very strong evidence that we lost those receptors due to acquisition of color vision. So, we don't need them anymore because the, whatever the function that they played is taken over by the color vision that we have, by trichromatic vision, like, for example, we have monkeys, the monkeys, new world monkeys that are living in the Americas. They have dichromatic vision. And they have a larger number of olfactory receptors. So, so there's some, there's a weird interaction between the visual and olfactory system. Yeah. So another interesting interaction, again anecdotal, is that, I mean the receptors actually in the eye and the receptors that the olfactory system is using, I mean, they are proteins which are on the surface of the cells, right?

ALEXEI: 06:15 And they interact with various, you know, stimuli. So they're very similar genetically. They're both built on the same type of receptor. And interestingly, the visual, you know, that olfactory receptors, they bind with different molecules and, you know, by the strength of the binding, by the affinity of this binding, they recognize what the molecule is. That's basically how it works. So some receptors don't bind certain types of molecules and that information or some, some other receptors will bind, uh, the molecules. And that information basically is used to derive what is the, what is the stimulus, right? Which receptors activate and which, which are not. So the same thing happens in visual system, because the visual system, the protein in the visual system actually relies on a small molecule, which is bound to that protein, which is derived from vitamin A. So that small molecule, so essentially, like, the visual system protein is, is an olfactory protein that binds another small molecule called retinal that changes conformation when the light, you know, when the light is coming in at certain wavelengths. So, and that's how we recognize different colors.

CRAIG: 07:27 That's interesting. The number of receptors in the human olfactory system, you were saying in a mouse,

ALEXEI: 07:36 It's about 1,100. It's about three times more.

CRAIG: 07:39 Right. So does that mean that the mouse can differentiate between many more molecules than.

ALEXEI: 07:45 Oh yeah.

CRAIG: 07:46 Yeah. And I've heard dogs as well.

ALEXEI: 07:48 Dogs actually have slightly fewer receptors, surprisingly, than the mouse. So it's surprising that dogs are, you know, actually hunting mice, rather than the other way around. But dogs have much larger [olfactory epithelium](https://en.wikipedia.org/wiki/Olfactory_epithelium), meaning that they have more cells basically. So probably the threshold for just detecting odorants is lower. So they're more sensitive in that way. And again, I mean, those systems are very strongly fine-tuned because those two animals, those two species, their existence depends on the sense of smell.

ALEXEI: 08:23 Right? Uh, and when I say that it's the most mysterious of senses, I'm being honest. We don't really understand what are the evolutionary pressures. That's what we're trying to understand actually through machine learning and other approaches, you know?

MUSIC 08:37 Musical Interlude

CRAIG: 08:45 Yeah. So the machine learning application that you've worked on, what kind of sensors are you using to detect the molecules?

ALEXEI: 08:57 That's the, that's the thing. So there's no - technology hasn't developed sensors which parallel in their sensitivity and discrimination, the real biological sensors. So we're not using any sensors. So, what we actually try to do is we try to predict human perceptions of various molecules and we're trying to understand what are, like, the evolutionary mechanisms or the evolutionary pressures on the, on the, on those 350 receptors that we have. Right. So, can we describe those pressures using machine learning language? So we don't use the measurements of, which are obtained from, uh, from sensors because the sensors are very, they're basically indiscriminate. The sensors that we have, the best ones, are not really good at discriminating odors.

ALEXEI: 09:44 So the type of data that we have, there are two types of data. There is data conformations of, of like, basically shapes of molecules. And I mean you might wonder how is it useful, right? Because it doesn't have any olfactory perception. So, we just have, like, molecules and how they exist, how they are shaped in 3D space. But it turns out that for machine learning, this is actually very useful information and I'll describe later how, right? But that's one of the types of data that we have. And this is virtually unlimited data set because we know everything about all the molecules, like, chemists are not wasting their time. They are computing. So we have about 10 million molecules in our database that we know everything about. We know conformations. We know their physical and chemical properties. So it's an enormous dataset. Unfortunately we don't know how they smell.

ALEXEI: 10:34 So we know how molecules smell for about 4,000 molecules. Right? But let me actually go back. So, so what we're trying to do here and what is really missing, the real question is the analog of the color space, like, in the visual system, we know that there are red, green and blue basically. Right? And we can synthesize similarly based on that knowledge using, like, a TV screen. Right? So, so the fact that we have such an efficient synthesis of visual stimuli is a testament to our knowledge, to our understanding of the color space. Yeah. That it's basically three colors, right? And we don't have a similar understanding in olfactory system, so we don't really know what are the features of odorants that olfactory system recognizes. And if there's any simplifying structure in that enormous space of molecules. I mean there are about 10 million volatile molecules.

ALEXEI: 11:37 Is there anything that we can like, do we simplify it or we just, like, basically recognize every molecule as a, as a new, unique, you know, entity. So that's the basic question. And so there is a hypothesis that we're trying to formulate that there is a simplifying structure. There is something, like, red, green and blue in olfaction. And basically, try to understand whether, what, we say, olfactory perceptual space is low dimensional, like, the color perceptual space is three-dimensional. So the question is if it's possible to describe the space of percepts of molecules, human percepts of molecules by a low dimensional, by a small number of parameters, right? So that's, that's the hypothesis of low dimensionality of olfaction. And it's interesting that to some degree you can, you can, like, take the, like, olfactory and perceptual data set and try to reduce dimensionality and a lot of information can be captured by low dimensional projections of those percepts.

ALEXEI: 12:43 And the interesting thing is what are those dimensions, right? So, like, what is, what are the analogs of red, green and blue? And we don't really know and to this point, like, how to attach an, an anecdotal meaning to them, although they do exist in the computer, we don't really understand what the meaning of them with a small number of exceptions. So one of the most interesting dimensions in olfaction turns out to be pleasantness of a molecule. Pleasantness. How pleasant the molecule is. This is something that like, for many molecules, people find universal, like, many molecules are uniquely pleasant or unpleasant. You know, that's, like, basically one of the most dominating dimensions. But, perhaps. If I were to bet, I would say that there are probably about 10 dimensions. Ten analogs of red, green and blue. And if that, if that were the case, if that was the hypothesis that would be confirmed by data, the, that will be a big deal, I think, because that will really drive the research.

ALEXEI: 13:47 Trying to understand what those dimensions are. Another question is, you can ask, is how do you describe olfactory percepts, right? So this is something that exists in our head, right? How do you put it out there? How do you put it in a dataset? like, when you smell a rose, how do you make, how you convert that into data? Right? This is actually a very imperfect process at this point. There are many different methods, like, for example, people could try to compare two molecules, how similar they are, right? And then probably if they are similar, more similar, then, that, you will find that they're closer in this perceptual space. I have, like, a term for this perceptual space. I call it smelliverse, you know, just basically olfactory color space. Right? You could try to find similarity between percepts for pairs of molecules and then to build the smelliverse from inside just based on their distance information.

ALEXEI: 14:44 And that, surprisingly, that strategy fails dramatically because, like, no people will ever agree on how to compare apples and, like, smells of apples and oranges. So if you do that, then the only thing that people can agree is how pleasant things are. Turns out what you'll see is the smelliverse is one dimensional. It will be basically the pleasantness will be the only dimension and the other dimensions will kind of average out. It will be, they will be washed out by this lack of consensus, if it makes sense. Yeah. So, so that turned out to be not a good strategy unless you're measuring, like, pleasantness of molecules. If you are, like, specifically interested in how pleasant the molecules are, that would be that strategy. So another method is so-called profiling and that probably generates the best data sets right now and that is describing molecules with words. So there are carefully selected hundreds of words. And then you basically ask humans to say which word applies to which molecule.

CRAIG: 15:48 One word per molecule, or which words?

ALEXEI: 15:51 Which words. I mean, normally, like, in the data sets that we use, they're, like, on average, two to three words per molecule. And there are hundreds of words, like, literally, right? Um, and those data sets are built by olfactory experts, so they kind of know where, where those words are located and how to describe molecules. So that is very imperfect type of data because it's semantic. Yeah, right. It relies on the ability of people to describe the indescribable percept that exists in their heads with the limited capacity of, you know, semantic capacity. So, and then the third type of data, which is now appearing, is actually the imaging from olfactory cortices of mice. So that would be actually a better descriptor, but we don't have much.

CRAIG: 16:47 imaging of?

ALEXEI: 16:48 just imaging of, like, mice would sniff a molecule and then you would basically install a microscope in the, in the [piriform cortex](https://en.wikipedia.org/wiki/Piriform_cortex), which is the olfactory cortex in the mouse. And you will image the activation pattern of cells in the piriform cortex and you would, that is basically, I think it's the ultimate measurement, so it's not clear how it's going to fail at this point. The way it fails at this point is that we don't have much data for many molecules so that we can train machine learning models.

CRAIG: 17:18 You could have analogous data from human cortex scans. No?

ALEXEI: 17:22 No, no. That's, that's a very good question. I mean you're asking a very good question. But the idea is that, why is it not possible comes from the way, how cortices are activated by smells. It turns out that essentially when you present the smell the cortex will respond in a white noise manner, so it will look, like, every small area is activated, contains activated cells and they're scattered around, like, pseudo randomly. The pattern is reproducible, meaning that if you present the same or, like, a similar smell, it's a similar set of patterns, a similar set of cells will be activated. Overlapping set of cells will be activated, but it's salt and pepper. So, you basically see, like, salt scattered all over the piriform cortex, right? So you need to have a very high spatial resolution to study that pattern. And, like, if you try to use methods that are available in humans such as [fMRI](https://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging), functional magnetic resonance imaging. So you will see their spatial resolution is not good enough really to resolve the subtleties of this, of that pattern.

CRAIG: 18:28 But in a mouse.

ALEXEI: 18:30 In the mouse, you can install a microscope near the cortex and see a small chunk of that activation pattern. But at high spatial resolution,.

CRAIG: 18:40 mic - literally a microscope.

ALEXEI: 18:41 literally a microscope, yes.

CRAIG: 18:43 And how does the activation visualize then?

ALEXEI: 18:47 There is a fluorescent dye that is activated by calcium. So, when the cell responds, some, like, calcium is released from the, you know, internal stores of the cells and that release of calcium activates a [fluorescent reporter](https://www.ncbi.nlm.nih.gov/pubmed/16988433) that, you know, fluoresces. And you can detect that, it's especially designed mouse, not a regular mouse.

CRAIG: 19:08 Yeah, that's fascinating. And so what do you do with this data set? And you don't have enough data

ALEXEI: 19:13 That data set, although available, it's not really useful at this point because there are literally, like, data for maybe, like, 20 smells and, like, machine learning models, they rely on, you know, thousands of data points to be, you know, useful, predictive. So, I think this is, like, the future, but at this point we don't have that data. We have data for humans.

CRAIG: 19:40 Yeah. And that's, that's the semantic piece. And you average that and you could come up with some sort of a, uh, it's not a high resolution, but you do come up.

ALEXEI: 19:52 Well, it's not actually so bad. I mean, it's not so bad because it's human, right? So that's what we're interested in. We're not really interested in mouse olfaction. Right. Although that's, that's interesting. On its own. Right. And, like, another thing which is interesting about olfactory system is that it's very similar between, like, vertebrates, invertebrates. It goes back to basically some very low-level organisms such as flies, like, flies, for example, fruit flies. They have a very similar olfactory system. They have very dissimilar visual system, but the anatomy, the structure of olfactory system are basically similar to humans. So which is fascinating. So once you understand one, so there is basically this notion that if you understand one, you can generalize this knowledge to, to another, you know,

MUSIC 20:42 Musical Interlude

CRAIG: 20:49 Using the semantic data that you've collected, how large is that data set?

ALEXEI: 20:55 So, we didn't collect it. We use the efforts of, you know, olfactory researchers. Uh, I mean we're computational people. We don't really collect data. So, so it's a very good question. So there are several data sets and they different qualities. So I think the best data set that is available right now is collected by company, which is called [Good Scents](http://www.thegoodscentscompany.com/). I mean, Good Scents. It's freely available for downloading. It's basically about 4,000 molecules, a little bit less than 4,000 molecules. They are projected to about 600 words space. So, and again, it's discreet, it's mostly what, zeros and a few ones, corresponding to the words that you know, describe that molecule. Uh, so it's basically a huge lookup table. And for every molecule, you know, some number of words varying from one to maybe eight to describe that molecule.

CRAIG: 21:53 But if you're training a neural net, I presume you're using deep learning on that data set. Then what are the inputs that, because you're working toward a classification system. So what are the inputs then? What are the classifications? Are you classifying specific molecules or families of molecules?

ALEXEI: 22:11 Yeah, we actually do two things. We are very good at one thing and then not so good, we're not perfect at another thing. And I'll describe what those things, what those things are. So how does deep learning apply to this? Does, where does it come in? Right. So, so the main hypothesis that we have is that olfactory receptor is basically this protein, is basically a spatial filter that interacts with molecules and extracts features from the molecules, certain chemical features, right? So, it's a spatial filter and you know that, like, the, like, in the visual domain, the deep learning networks, they are very successful in classifying images, right? So our idea is to, basically this protein is a, is a network, is a network of the means, the network of different parts that interact with each other. I guess network is another way of describing interactions, right? So the olfactory receptor is a network that describes the filter, that interacts with molecules, it extracts certain molecular features. And that's how we can use deep learning. That is our hypothesis. So that's the overarching hypothesis. And we use it in two different applications. And one application is to describe the space of molecules, for all molecules in the universe, which are volatile. And how would you do it? And to do that, we do something which is called, which is called [autoencoder](https://en.wikipedia.org/wiki/Autoencoder).

ALEXEI: 23:55 So the idea is that you present olfactory receptor that is a neural network very similar to the visual, except that it's three dimensional. And so you present this molecule to the network. And what it does, it basically converts it into some internal representation, right? Which we argue corresponds to the olfactory to the 350 olfactory receptors. Right? And then you have another network which takes that internal representation, which are the 350 olfactory receptors, and expands them back to the real space, three-dimensional space. You give it the molecule and what you get on the exit is, is the same molecule. So you train the network actually to replicate the molecule itself. And how's it useful? It looks, like, highly redundant, right? But the most important feature is that we, the middle layer, which is the olfactory receptor layer where you have the latent features, that code that molecule are, it's severely compressed.

ALEXEI: 24:51 So the amount of information in that layer is very small. So the idea is that can you put the molecule, can you compress the molecule to such a degree that basically extracts the most important features in that molecule and molecule is still replicable, replicatable on the output of the network. Sure. This is a common technique.

CRAIG: 25:14 it's encoding?

ALEXEI: 25:17 Encoding and then decoding, right? But then in the middle there is a bottleneck. There's a tiny hole through which all information about that molecule can come in. And if you are successful in replicating the molecule using this minimal information in the middle, then you have this very compact and efficient representation of the overall olfactory space. Right? So why is it useful? And this is useful because in the visual system when there wasn't enough classified data, so this basically relies on the unlimited data set that we have that I described of, of structures of molecules.

ALEXEI: 25:53 It doesn't require any olfactory percepts, right? You don't need to know how the molecules smells to train this type of autoencoder and therefore we can rely on the millions of molecules that we have structure for to basically to pre train this network for a more challenging task, which is understanding the olfactory process. So it's basically a pre-training step, but it's also interesting. It's fascinating to know that you can compress this enormous molecular space and put it through the small hole in the middle. We estimate that you basically need about forty receptors actually to reliably replicate the molecule to a very high precision.

CRAIG: 26:34 The end goal though, as you said, you want to be able to classify molecules according to their smell. So you can go through this process with molecules whose smell you can describe using this semantic database, right? And then you can present molecules that have not been captured by that semantic database and estimate where within the smelliverse they exist. That's fascinating.

ALEXEI: 27:03 So, so this is, like, the first, the autoencoder is the first step. Perhaps I'm a little bit too technical but, but describing all the steps. But this is basically something that we're really proud of because it works, right? So, we put the molecules, we, like, shred them and put them through, compress them into a tiny hole in the middle of latent representation. And then we expand them and we get the three dimensional shape. Perfectly, right? So that is great. But then the second step is actually to use that information, that latent information to understand how molecules smell. So what you do is you take the part which encodes the molecule, which shrinks them to this small, to the small, through this small hole and add, add a classifier, another deep network, which is a classifier that will then convert that latent representation into the percept. Right? So that works, like, fairly well, but it's not really perfect. I think there is a little bit of, there's a lot of actually room for improvement. So, you know, we for example, can predict pleasantness of molecules with a correlation of about 0.7 so we're about, like, 30% away from the perfect predictor

CRAIG: 28:15 just on pleasantness. Right. But presumably then you have all of these words that are being used in, in your data will eventually be able to present them all to.

ALEXEI: 28:25 We have to get about to 0.9 to be able to predict, predict individual words

CRAIG: 28:30 Like, citrus or I was looking at some of your data. What was the one for decaying, uh, or cadaver

ALEXEI: 28:39 Cadaverous, words, like, that. Yeah. The ultimate goal is obviously to basically put the whole 10 million smell, ten million molecule database and to predict individual words for every molecule. That would be basically what we're trying to deal with. So, we estimate that our, like, predictive score has to be about 0.9 for that to be possible. So, at this point we have about, like, 0.7 but we didn't really exhaust all the possibilities for us to improve it, improve the classifier and the autoencoder is one of the most important steps I think. Um, in terms of its, like, one of those bells and whistles that is very important because we're dealing with a very limited data set of about 4,000 molecules which, we know how they smell. And I think, like, for image classification you need, like, millions of images, right, to build a predictive network. So we are at the point, at the very early stage when we don't have enough data to be really predictive.

CRAIG: 29:50 Yeah. What are some of the applications? If you reach a point where you have very high accuracy probability and then you start feeding molecules beyond the 4,000 through the network. So you're classifying the smells. How do you use that? I'm sure you've thought about applications, how this can be used.

ALEXEI: 30:13 I mean it will give us an understanding of the sense of smell, right? So, we'll have a clear picture of what are the features which are extracted from the molecules and that has, like, fundamental significance for humankind, because we don't know how one of our senses works. Well, we know some of it. We know some molecular mechanisms, but we don't really understand on the same level as you know, [Helmholtz](https://en.wikipedia.org/wiki/Hermann_von_Helmholtz) and [Young](https://en.wikipedia.org/wiki/Thomas_Young_(scientist)) understood, like, early in the 19th century. Yeah. So, we are basically, like, 200 years, we're delayed by 200 years compared to the visual system. Right. So which is ridiculous, right? So 21st century we don't understand how one of our senses work. So that's not an application. I think that's basically one of the fundamental problems that people face, although they're probably not aware about it. Right. But in terms of applications, of course you will be able to generate molecules with a given smell, which will revolutionize perfume industry.

CRAIG: 31:13 Sure. I would assume there's a correlation between the pleasantness or, or what's the opposite in your unpleasantness? of a smell with its health effects. For example,

ALEXEI: 31:26 there's another aspect which is the uh, the uh, basically predicting, not necessarily olfactory percepts but for example, the, the toxic effect, the toxicity of different, uh, that's what you're alluding to.

CRAIG: 31:40 that's what I'm trying to say.

ALEXEI: 31:41 Well there are also, like, data sets for that. We could use, you know, the same network to, to predict toxicity for example, or how allergenic given compounds are and that would be also important. Obviously, it's an important application.

CRAIG: 32:00 The receptors. When the molecule binds to the receptor, is it physically fitting? Is it the shape of the molecule that allows it to bind?

ALEXEI: 32:10 Well, that's another fascinating question that nobody knows the answer to. So there are two dominating theories and one of them is steric theory. Another one is vibration theory. So, I think steric theory is mostly accepted by, like, most of the community. But so, the steric theory of ideas, like, exactly, like, what you said, that you have, like, a key and there is a key hole, right? Molecule is the key and the receptor is the keyhole. If they fit then the receptor is activated, right? So, it's basically, like, shape matching, right? I think if I were to bet, I would give that my preference, that explanation. There's an alternative theory, which is vibration theory is that molecules have a vibration spectrum, like, basically, like, bending of chemical bonds and that vibration spectrum is somehow recognized by olfactory receptor, not specifically the shape. Of course the vibration spectrum depends on shape. Yeah. But, like, molecules with different shapes and the same vibration spectrum could be, sensed as similar molecules. Right? So that's the hypothesis. So there is some evidence for vibration theory, but you know, it's, it's questioned by most of the scientists. So that, that is one of the things which is not really understood.

CRAIG: 33:23 And when you have the semantic database, any particular molecule, there's a range of words that people apply to that molecule. Is it the case that there are multiple receptors activated by one molecule? That may be, there are some receptors where one part of the molecule fits and then other receptors where the other part of the molecule fits so you get multiple,

ALEXEI: 33:50 Right. That that is what is called combinatorial code. So, the understanding is that, like, the same molecule will activate a large subset of receptors, perhaps somewhere from 10% to 50%. It depends also on concentration, like, higher concentrations will activate larger number of receptors, like, generally. And then this, the pattern of activation, so take those 350 receptors, will convey the odor identity and concentration. Right. So unfortunately I cannot be more specific. It's basically unknown. Yeah. What is, what is the logic of that? Which is not really sparse. Right? So you have maybe from 10 to 50% receptors activated. Now the interesting question is that, like, the same molecule of different concentrations is perceived as similarly. Right. And what does it similar, what is similar between those patterns for different concentrations that makes the molecule appear as a single percept? Right, like, for example, when you try to find a coffee shop, right? You follow the smell of the gradient of coffee, of smell, of coffee, right. And, and the smell of coffee, it will be similar both, like, a hundred meters from the coffee shop and, like, 10, 10 meters from the coffee shop. Although the pattern of activity of those receptors will be quite different. Uh, there'll be more receptors activated in the latter case. Yeah. So what is it? What is it? How is it that we interpret it as the same thing as the smell of coffee? There are some speculations about, some theories about how this happens.

CRAIG: 35:23 So going back to where you are in the research so far on the classification and you said you're 70%.

ALEXEI: 35:30 our score is about 0.7, which is, like, this correlation coefficient that describes how well we can, like, describe pleasantness of a molecule.

CRAIG: 35:40 Yeah, just, just pleasantness. Yeah. To get to the individual semantic labels is, is it a matter of you need a larger data set or it's a matter of refining the,

ALEXEI: 35:56 The existing, the model, the network? Well I think both, like, obviously we can, uh, improve our models and we have some ideas of how to do it, but I think ultimately both are necessary. So, I think we, we have some tricks, you know, in our sleeve that we are prepared to use in the next, like, a few months. But also, we have to have probably more data because, like, 4,000 molecules is a little bit difficult. Only experience will tell, of course, like, it's much easier to introduce bells and whistles into deep network and it's much more fun than dealing with collecting experimental data. Right. So, I anticipate that we can basically improve our networks to as much as possible.

MUSIC: 36:44 Musical Interlude

CRAIG: 36:53 Are there sensors that can identify molecules now in the air?

ALEXEI: 37:00 Uh,

CRAIG: 37:00 I mean I, I was in Iraq for example, and they gave us these stickers that you put on the outside of your vehicle that was supposed to detect chemical weapons. So.

ALEXEI: 37:10 I think, like, mass spectroscopy would be, I mean it's very expensive and heavy equipment, but it can do it. It's the one which is used in the airports to detect ...

CRAIG: 37:18 used in airports for?

ALEXEI: 37:19 for detecting, like, explosives when they swab you. But again, like, the thing is that, you know, the mass of the compounds - smells that we're exposed to, like, the smell of coffee would include hundreds of components or, like, a smell of lemon, like, apple.

CRAIG: 37:33 components, meaning molecules.

ALEXEI: 37:34 individual molecules. That's another interesting and fascinating question is how they're combined, like, if you use mass spectrometer to identify components, you only know their mass. So you have to basically then resolve and understand what the individual molecules are. At this point, artificial noses are not really a viable alternative to basically biological, biological systems. And my understanding is that the resolution, resolution in terms of, like, identifying different molecules is not good enough to match, like, even the human nose.

CRAIG: 38:10 So what's next in your research?

ALEXEI: 38:12 So basically the next few months we were going to, like, add bells and whistles, like, more bells and whistles, like, explore other deep learning methods. Then another interesting question is basically how to introduce mixtures because real smells are mixtures, so that would be also fascinating thing to add to the architecture of the network.

CRAIG: 38:36 Oh, yeah. Yeah. Wow. That's fascinating.

CRAIG: 38:41 That’s it for this week’s podcast. I want to thank Alexei for his time and hope everyone listening has a renewed appreciation for their olfactory receptors and the volatile molecules that excite them. If you want to go into greater depth about the things we talked about today, you can find a transcript of this show in the program notes. I encourage you to download and read because I find a transcript captures and clarifies details that the ear misses. You can help us by rating or reviewing the podcast on whatever platform you use to listen. It helps increase our visibility and reach more listeners.

The singularity may not be near, but AI is about to change your world. So, pay attention.