# Where ChatGPT gets its products | Tom Wells, Researcher @ Peec AI

Masters of Search · Episode 34 · Published 2026-03-24 · Length 1:03:13 · Host: Niklas Buschner

Source: https://www.masters-of-search.com/podcast/where-chatgpt-gets-its-products-tom-wells-researcher-peec-ai

Listen: Apple Podcasts · YouTube · Podigee

## Show notes

My guest in this episode is Tom Wells, AI Search Researcher at Peec AI. Tom has spent over a decade decoding ranking signals at companies like Semrush and Searchmetrics, and he recently published some fascinating research that got a lot of attention in the ecommerce SEO and AI Search world.

Check out Tom's research:

- https://searchengineland.com/new-finding-chatgpt-sources-83-of-its-carousel-products-from-google-shopping-via-shopping-query-fan-outs-470723
- https://peec.ai/blog/how-ecommerce-managers-can-optimize-chatgpt-product-rankings-step-by-step

Let's connect! Niklas on LinkedIn: https://www.linkedin.com/in/niklas-buschner/ Radyant on LinkedIn: https://www.linkedin.com/company/radyant/ Tom on LinkedIn: https://www.linkedin.com/in/tom-loves-data/ Peec AI on LinkedIn: https://www.linkedin.com/company/peec-ai/

## Full conversation

**[00:00:00]** We do have some innate instincts to actually access what is interesting to us as a business. Can you ask or find interesting questions that you want to solve? Because if yes, You really need step one mailed down for great research and can talk about the key findings in this study without spoiling too much. What was it that you found? Yeah so Confirmed what I heavily suspected that the vast majority of products, That make their way into the chat GPT product carousel are in fact sourced from Google organic shopping results. And for me this is like a scene In The Wizard Of Oz where someone peers behind A curtain and you see it's not actually magic anymore. Before we dive-in You're listening to the Masters Of Search podcast with your host Niklas Buschner.

**[00:00:52]** Each week, I sit down with some of these modest people around the world in SEO and AI search to bring you their strategies mental models. And top pieces of actionable advice. if You enjoy this podcast don't forget to like and subscribe. then follow it In your favorite podcasting app or YouTube. It helps us get top-notch guests and create the best possible content for you. Let's dive into todays episode. my guest Today is Tom Welz AI Search researcher at PKI. Tom has spent over a decade decoding ranking signals at companies like Samrush and Searchmetrics. And he recently published some fascinating research that got a lot of attention in the e-commerce SEO, and AI search world. Excited to dive into this so thanks for coming on the podcast Tom!

**[00:01:38]** Thanks for inviting me Nicholas super happy to be here. Thanks for taking the time Tom. Let's do a quick intro For everybody that does not know you yet. So who are you and what do you do at

pki? And how did you end up in this whole like AI search SEO space? so yeah, actually my story goes all The way back to search metrics which believe it or not used To be arrival of Sam rush. when I tell people there's they're Like i don't know the name Or maybe I heard It and it was kind Of famous ten years ago. But then What happened? Which is fair enough, but believe it or not we were rivals with Bright Edge and Semrush. And actually doing pretty well! How I started there was that i entered the marketing department in two thousand and fifteen... ...and what I really loved.

**[00:02:23]** this sounds quite funny.. ..but the actual office space. so It's open plan. everybody could speak to every other team on same floor.... ....I remember being really fascinated by the data science team and everything they are doing. I noticed one of key gaps at that time, ten years ago was the divide between marketing departments who were focused on e-mail and data science. I felt this channel in between two wasn't as good And this led to doing things like conference presentations, famous studies. Like the ranking factor study back then which was pretty well downloaded. and Then I just increased my data science knowledge from there leading on To things like doing a master of Data Science degree From Georgia Tech in really kind of obsessing with The data side of stuff having started more In the marketing site.

**[00:03:29]** So yeah i have both elements of it and I kind of love both elements off that world. And from then on went onto to do consulting, yes still in SEO as well but also in AI pipelines building them from scratch stuff like this. returning a while ago to SAMRAS particularly with the interest in AI search going there when Perplexity launched doing Seeing the opportunity to move across to PKI and what really inspired me to make The Move, or you could call it Make the Jump from a thousand-person organization where I'm doing very high level research with many resources at my disposal. To peak was exactly the opposite now i am responsible for effectively running the Research Unit as good as possible With slightly fewer resources but actually found that quite an empowering idea.

**[00:04:25]** basically. so yeah thats why im there. Sounds pretty much like the motivation from Malta when he decided to leave it yellow, right? And then go to PKI. Did you also have a chance to work with him back there at Search Matrix or did you two overlap? We did overlap especially in the twenty fifteen-twenty sixteen period so already ancient history. now Malta was trying to bridge between data science and marketing particularly for primary primary research, so I helped him there. And we also reconnected in a funny way when he at EDILO went effectively on a mission to try and prove that Google Shopping was in someway manipulating the ecosystem? There is famous lawsuit where EDILOS sued Google successfully pieces of primary research.

**[00:05:21]** I was actually the author of this for Malta, so while i were still at search metrics ,I did a study looking at basically Google shopping vendors and proving that the majority off them are basically still somehow with google entity or some how influencing The Google rankings. then This is part of her successful lawsuit against Google. So PKI is doing a pretty good job in pulling together. A lot of brain power from the whole SEO space and like OG SEOs for people that might also be interested in, this whole research thing because I see more companies investing resources into research publishing their own studies etc. can you just take us through normal day super regular days off your job as a researcher at PKI?

**[00:06:13]** Absolutely. I mean, if i could say there is a regular day of PKI right now we're in a scale growth phase so the days are very different. but maybe I would just preface this response. So people in the audience know why it's actually... Why research department is even a thing In such small start-up? because you can build apps and execute Now quickly with AI coding software And come up pretty good stuff and get it on the market very, very quickly. However many of them don't even have a research department and never even considered this. for me This is kind of crazy because I always say in the

world of AI search we're in the business Of understanding how AI Search works And this is impossible In a probabilistic system without actually researching How It Works right?

**[00:07:05]** We can do The basic level throw in some prompts and see the responses, but does this actually tell you how ChatGPT works under the hood? I would answer probably no. And so having this quality primary research of how AI search works across all threads is critical for business understanding, product understanding, potentially educating industry as well. So that's why even a small scale startup having a good research unit, even if it's one or two people can make huge difference to what you're actually able to achieve basically. A normal day. so we split our research into quite few different buckets. There can be short-term trends where we have to basically react to something as now. a couple of weeks ago, one of the big things on LinkedIn was that chat GPT query fanouts were announced.

**[00:08:04]** That they disappeared. So I spent then three days researching this. two actually gets do The truth? and the answer is it's nuanced more complicated than than what some people claimed on LinkedIn, let's say. And I published that in a response to give more detailed opinion of how this is reactive. Then we have long-term research which i'm sure it one other things will talk about today where an idea can become An obsession or request To prove or disprove something and It could take anywhere from one or two weeks even four months In the case for shopping studies. And of course, we collaborate with some of our great clients come up with actually amazing research ideas. So instead I've just show them the door and say hey!

**[00:08:53]** We are the ones with all their knowledge... ...we have an open-door policy to people who want to collaborate with us.. ..and i evaluate every idea as if it was brought me by the same person. there's no hierarchy that. so um we have a few research pieces coming up. they were actually inspired by our client said Hey could you actually Check out this? wouldn't this be nice to know the answer to this question? and then we decide internally. Hey, can we actually do it? It would be really cool if we could. um And that's more or less how-how we work at the moment. Awesome! If there is someone in the audience That also fancied Starting their own research in any way, like let's even say we for example at radian So we already thought about hey.

**[00:09:37]** How can we do some cool research? For example around time spoiling an idea here Some cool research around how customized are really the responses from chachivity or other AI search interfaces Depending on the persona that we are. so were simulating different personas Like Let's Say a CEO Or A CMO an SEO manager. And if so, If we think that this will be loaded from the memory of chat GBT how different are the responses? This is something that we for example just thought about. Hey, this would be cool to have some solid takes on. How would you recommend starting with research if someone sees The value and says okay yeah Tom You convinced me. how do I start like today? That's a great question, Niklas. I really actually like this question because what you touched on is actually what i would call internal business intelligence but also combined with a bit of instinct.

**[00:10:33]** so your instinct that this might be interesting? I would say ninety percent of the time for clients. it's actually correct that they have this question that they've kind of been obsessing with for months and its being at the back their mind. then they didn't know how to surface as a question or well formed research study. And I do find it interesting that we have some innate instincts to actually access what is interesting for us as a business. Yeah, i think this kind of the first step. can you ask or find interesting questions? because if yes? You really need to step one and nail down on great research from there... This gets more complicated where might not be so easy. but If you ask the question, one really good thing is to basically do what I call which is effectively design.

**[00:11:27]** The optimum version of this study that you would like to see at the end and then work

backwards from here. so in your case i would say okay personas for Radiant. maybe we have these five personas That We know In our ICP ,we can actually describe them Really well. Our Ideal Study Would Be Very Detailed At Giving Information On All Of These. How do we then access this information? So in the prompt universe, you know This is one of the exercises that any AI search company is able to Do as we translate. The idea for study from the abstract into prompts. so We would say okay Maybe before this persona they were asked prompts In a different way on the back end. Then we can measure the difference of the information.

**[00:12:12]** Of course there are A lot of different technical skills That I'm not going To go too deep Into now. however i Would Say the core part Is asking a great, interesting question and sort of even imagining like how could we get this data back? And then that is already at the stage where you can give it to a developer or researcher. Or someone they could actually deliver out for you. awesome I think That's very helpful! And also see my idea now being validated. but honestly i have to give shout-out To our internal AI search strategy lead Yannick who came up with The Idea because We were fascinated by Already put out and we wanted to do something similar around things that We just felt is interesting. And obviously, we also feel like we can't Just offload all our ideas on you because You only have so much time in the day.

**[00:13:01]** But let's talk about Some of the research that you actually published because I already mentioned in the intro, it's about eCommerce SEO and AI search. For everybody listening we will also put all relevant links on individual description or show notes so you can check very extensive write-ups from Tom and team. give us a quick like very brief intro. what was your idea with this research? Like, how did you also come up with the idea? Yeah. So in October last year this is when it started I noticed that ChachiPT was... It seemed to me leaning on Google Shopping results so it didn't actually start as a fully formed idea. more interesting is maybe how i notice this. there were very few researchers one of them being Olivier de Segonzac from Rezonio in France, me and him had some messages with each other about hey are you seeing this?

**[00:14:06]** why does no one care? And the trick for me was to see that I just did a random sample of literally on my personal chat gpt account asking questions about e-commerce products. And then I thought, well now think about it a bit more carefully. There's no way that Google would give open access to their shopping graph to chatGPT who could be considered arrival at this stage. Then i decided the only logical conclusion is they are scraping these results via third party scraper. So for people in audience you don't know really web scraping its effectively where You pay your provider emulated as if you were a real person. So, I want the Google shopping results for best running shoes under five hundred dollars in an area of the US.

**[00:15:06]** let's say You can then pay a scraping provider like SerpAPI or search api.io to name just two and Then effectively get those results back. now kind of a complicated explanation into the study, but it's as simple I could frame. It started with belief that chatGPT was scraping Google Shopping which is already very big finding. if it would prove to be true Nice! Can you talk about key findings in this study without... Spoilering like too much because we will go into the details one by one, but maybe high level. What was it that you found? Yeah so this study confirmed what I heavily suspected. That's the vast majority of products that make their way to the chat GPT product carousel We've probably already seen are in fact sourced from Google organic shopping results.

**[00:16:05]** And it makes me happy to say this because the journey to prove this, to be true took me three months or four months of work essentially. I still see a lot people claiming that Chatchapiti sources its information if they do web search and grounding from Bing. So did you also check? maybe this comes from Google but maybe it's just a coincidence. so did you think about that? So actually a shout

out here to the Search Engine Land editorial team. Part of the long process for getting published on reputable journal or website is that editors have to recreate this study with their data science team, at least take it apart and look from conceptual point-of view like they will not publish something they truly don't believe when its headline claim like chat you pt sources.

**[00:17:00]** And they actually came up with the idea when I first spoke to them early in January that it would be really good to add a negative control and experiment. It's exactly what you say, Niclas! That is not random chance for viewers thinking like yeah but it'd be easy to prove no its not trivial because If you imagine a very wide set of products like running shoes, watches with hard rate monitors or any e-commerce product that could possibly be imagined effectively we want to diverse the source as possible. so it's true across all them. You might assume probably top brands would appear on both. It turns out that's actually not true, but that was also not obvious. So we ran the exact same data pipeline on both Google Shopping Organic and Bing Shopping organic.

**[00:18:05]** so yeah... We did do the negative control And while we found about ten percent of products matching the ChachiBT carousel from being of those matched Very, very few were not also found in Google with the overlap. So we were able to sufficiently prove that no they are not using being as a source. basically and can you talk a little bit about the methodology? because I mean people would probably think okay but is this true across various verticals? so did they really check? A couple of different product categories or is it maybe bias in certain categories? how do look at the matching like? what does it mean that uh, The vast majority of products are matched. So how can I actually think about this? so Can you just share a little bit About How You Approached It Maybe Without Being Too Technical Although I Think The Audience Is The Audience is Probably They're-It's Very Intelligent People i know That!

**[00:19:09]** So um Yeah...You Can Go Down A Little Bit Into The Rebitall yeah..so Yeah, as you say without going into the rabbit hole whenever you do a study and I would encourage any Research team to sort of maybe follow this one or two pieces over advice. It's not me preaching it Just if they already good at research probably then know this already. But if you're just getting started into it This is a really important thing that When you first create let's say in this case The product set did you want to check? Not only do you want it to be very diverse across categories. You want what's called data harmonization. So, ideally one a similar number of things in a similar Number of categories across the symbol is similar number like this.

**[00:19:53]** so then don't have some Let's say running shoe bias and the data. if we have twenty thousand queries about running shoes My findings wouldn't be very good right? Like this is obvious. so We had somewhere in the region off a hundred subcategories of products and a total of forty-three thousand unique products that we checked in ChowGBT, two hundred fifty thousand products. In Google Organics Shopping And two hundred and fifty thousand product in Bing Shopping. Okay! How did you look at the match? Like because I saw on their search engine land piece That actually have very sophisticated way of thinking about the matching and I even so honestly, i think generally understand what you did there but just on a very distant level.

**[00:20:46]** So maybe you can share little bit about that? So firstly like why in search engine land we were so robust with the methodology...I will share this completely openly with all the viewers is because im employed by PKI as researcher. The chance of me being able to cherry-pick or influence the data and publish it on search engine land is more than your average researcher. However, I take my job very seriously as an absolute neutral person And as i said this obsession began before my job at PKI already in November. So we published a lot. methodology that we appreciated in advance might not fully understand but it was important for me to have it there, so also someone could recreate this study if

they had the ability.

**[00:21:37]** The ideal case would be all of their product titles from ChatGPT exactly matched Google Shopping. however Google Shoping and ChatGpT for that matter have the ability to dynamically rewrite small amounts. So if you do a test, You can ask the same question to Google Shopping one hundred times and you'll see slight variations in the titles that are retrieved. Sometimes it might leave certain aspects of... If your looking for smart TVs It may include dimensions already on product title or not. However the dimensions don't actually count as match because its not with brand. So to create a way to actually efficiently get past this dynamic process, because we're dealing with two probabilistic systems and we must somehow try and match them.

**[00:22:36]** I created a three-step algorithm to do this And when people hear the word Algorithm they got a bit scared. However it's just a repetitive process that works across the whole dataset. so step one exact match very easy. So if the strings or text match exactly, it's a hit. This is an easy case! Then we did couple of more advanced things so basically said okay what about if word order is mixed but words are same? If you imagine a silly example would be iPhone XI and we had iPhone XV. for whatever reason this wouldn't come in the data. But sometimes when they're switched still counts as a match And as a third, we actually counted the number of similar tokens it's called by subdividing the words into different parts and then match there.

**[00:23:32]** There was also random sampling manual controlling many other controls in place... ...and effectively. with all this we decided that we want to set a threshold for matching where the brand or product is the same. So basically we set our threshold at zero point eight. That allows for those things where there's a slight deviation in the title, but we're still very confident that the brand and products are the same. And this is why I can then say as the results of study eighty three percent make this threshold interestingly enough already above. forty percent made it into the exact match category which has also huge finding itself. so summarizing this in like Very simple way you were saying based on forty five percent of products that show up in chat GPT.

**[00:24:26]** That are exactly the same in Google organic. this is such strong evidence with also then more than thirty seven percent off products being almost an exact match, but there's no other reasonable explanation that these products come from the organic shopping results for google instead correct? Feeling like a lawyer here, in court. This is correct from the interpretive side. just to give you more details on why... In this case it's very rare that we come up with these confident claims unless were sure The real reason? going back already to November if you remember I said at top of show i looked and basically saw prices and ratings seem be same. Then I went a step deeper and looked at the source code of ChatGPT to see what is happening there.

**[00:25:28]** Like, What's it searching for? How are you thinking about these products? And as stated in search engine land piece There was a field Which has product token. When you decoded this fields Magically You saw what looks exactly like Google shopping parameters. Now Using these parameters, I could then reconstruct the exact URL that would link to a product on Google Shopping every time. A hundred percent of the time. Now this still does not prove what i wanted to prove in the study. This just says... decide who makes it first, second third fourth in the carousel. It didn't prove any of this and it also did improve at a scale where basically international research community would accept that however And still remains to this day.

**[00:26:26]** That is in the source code. everybody can actually look out and get Google Shopping URL from within chat gpc's source code. Now two questions. First do you think um, matters to open eye in any way. Like is it something that they would have. rather I know this was subject to interpretation. if you're not happy too share interpretation totally fine but Is there's some thing that they feel like? ah we

are not so happy with this now being public. and then secondly did you already get a letter from Open AI lawyers? Amazingly i didn't got anything in the PKI inbox And honestly I take it as a compliment to the methodology of research because, using steps in Search Engine Land article if you were out today get forty thousand products on chat GPT carousels and run them through using the exact shopping query finance which is something we'll talk about later.

**[00:27:27]** And then map that all back You would arrive at similar results. So...I don't think OpenAI have time my research, in the sense that they have a lot of other things too. However it is fitting in with the history of when...when they develop very fast and push features. I have seen in other areas this overreliance on the Google method. so for example um i haven't checked for a while if it's still the case but certainly things near you. They actually left Google Maps scraping parameters hidden in the source code as well. and it's more symptomatic of, we can ask like yeah but why would they do this? Don't they have The Greatest Engineers In The World And Raise The Biggest Fundraising Round In The world? Like Do They Really Need To Do This?

**[00:28:20]** Well I Would Say That It Took Google ten or fifteen, twenty years to make Google Shopping Graph what it is today. It's one of the most difficult feats of engineering. you know taking entities live pricing putting it all together and still has its flaws. for Chachi PT to recreate that in a few months would be very challenging I think. Got it! Do you feel like this supporting argument for the people that sometimes make the claim that AEO, GEO AI search is basically just SEO? I would say it's very different in a different AI models. Gemini Google AI mode potentially has the ability to pull directly from shopping graphs. so as an example although i was able to match all of these products if you actually went into the carousel and then checked Because it pulls also the image that is from The Third Party Scraper, you can then see basically deficiencies in chat GPT where if actually go through and try to buy a carry-on suitcase for your next business trip.

**[00:29:36]** And you want it black. There are cases of click on the ChatGPT link. It might be pink or out of stock. Why? Latency scraping latency So often To save cost. Any scraping provider will tell you that if, basically pay to have the thing scraped overnight it would be much cheaper when collected in the morning. But might a day late. so I also tested this across Black Friday and chat. GPT carousel could not keep up with live pricing data of Google Shopping. So sometimes differences are also in price. A scraping layer is always worse than real first party actual shopping thing. So that's just basically from a user perspective, however what AI can do I believe is the reason users like it. they summarize information and make into buying guide in style you would.

**[00:30:39]** It turns out that it's actually the combination of, you know your brand and product sentiment how you mentioned in third-party sources mapped In with how we appear on shopping feeds like Google Shopping And taking all together Of You Know How Your Overall Ecom Presence Would Be In AI Search Which Of Course Is More Complicated Than It Used To be. So I would say That If Anything The Discipline Has Become more Complicated i Speak With SEO Veterans Of Ten Years Every Day For work and you know, they are struggling stills to understand. And it's quite right in quite easy to understand that things have moved beyond the template links on the keywords. now It is a much more complicated system operating I would say.

**[00:31:23]** Talking about fanarts, you already mentioned Fanart queries. Let's look a little bit about how Chatchity actually pulls in the information. so How do they go from? A user prompt to the actual product carousel and everything. that I see as a response because You also found something very interesting there. not only That The vast majority of products come from Google organic shopping but There is more to this whole process. right Absolutely. So, you know fanouts has been a much discussed term in the online community and effectively You can just think of it as The type of web search that the AI does

to get the information That you want. I think this is A fairly simple explanation. We looked at query Fanouts in a lot of detail at peak over the last few months, and around November time.

**[00:32:18]** The same I was interested with this shopping topic. There's actually a doubling of average word length for search query. This is surprising finding to many people. They thought why would chatGPT suddenly make these fanouts very long? The reason that longer the Fanout Query typically more long tail or specific websearch results will be and potentially it's not always true, closer to what the user has actually asked. And basically chatGPT learned that if you ask a longer question You get more specific answers. This is part one of the answer. Part two of the Answer Is once you have More Specific results related To The User Query It's Actually Easier for the AI to differentiate between them and pick the right ones.

**[00:33:07]** Do you answer the? To answer the Question this is on the level Of Normal search fanouts, this is one of the trends happening there. Now in the field that I decoded that i mentioned where it was what's called base six d four decoded just means its not human readable. you have to use something to basically convert back into something you can read. within these fields was specific query the chat GPT using to retrieve products from Google Shopping encoded and still today both on five point four and five point three, this still applies. Now I find that's really interesting because if you just check Google Shopping with the same prompt that they use it to your accuracy would be very low. one can you target?

**[00:33:51]** The same products. but within these fields not only does try GPT encode the location that uses to search for the product which is people are interested in field UULE And If You Decode This It Gives You Locations So New York or whatever. If you combine the location with the actual query that it uses, the accuracy of what products ChatGPT is retrieving suddenly increases greatly. And this is how you can then do a study like I showed. So effectively chat GPT source code tells us exactly What? It's doing in the background. and so specifically about these shopping fanouts what we can say Is they are much shorter unless specifically targeting product categories. Whereas in a contextual normal search fan out, it might ask something like what do I need to consider when buying running shoes?

**[00:34:47]** Some long query where its targeting articles that can take chance of text and so the AI itself can understand oh you need other shoes waterproof are training for a marathon. this is basically what it uses to pat that text but specifically pulling products attacking a specific product category. Well, something I'm wondering is that JetGBT also introduced their own product feed? I think already quite some time ago and so why do you think or do have any information like did in anyway see this product feed? the JetGBD has launched themselves. How products are recommended? Is this something that is only relevant to Chatchapiti ads. So what's going on with the whole Chachapiti product feed?". Yeah, so data for the study was pulled in around mid-to end of January and back then a couple months ago now I did not see ChachAPiti's own feed be an irrelevant source across the forty three thousand products because if i had it would have been able eighty three percent of them or whatever the exact figure was.

**[00:36:07]** So what's happening there is if you think of retailers like Etsy, There's an increased move for Chachi BT to basically onboard big brands feeds such as Dick Sporting Goods Best Buy this type of things where instead of doing Google scraping search methodology it would then in the background try and retrieve directly from from other providers, so you have their own effectively data set or their own shopping graph if you like. The added complexity of this becomes... From a simple user query how would chatGPT decide which of those networks to pull the data form? So as often now as you know ChatGPt is still onboarding these providers and doing its testing. You see some volatility in the data while it's doing this testing And whilst on boarding.

**[00:36:59]** however If you look at You know, it was announced recently that ChatGPT decided to move

the Instant Checkout feed outside of this. This is one case where they pushed very quickly and basically do this but ended up not really working well. And now with onboarding different clients in their own product feeds there's quite a technical challenge for them spends infrastructure resources on actually somehow integrating. And this is also very, very tricky because in a query it kind of introduces the type of bias where chatGPT says well we have these ten providers you know? You can buy your running shoes from Google Shopping or Dick's Sporting Goods. however if you think about google shopping Is It A Democracy?

**[00:37:55]** Not Really But Its More Of A Democracy Than Going Straight To Dick's Sports Goods. Because let's say the most optimized fees, whatever that means are the products that appear on top and you do at least get a variety of providers or retailers who you could buy from. If you then go down the road saying well ChatGBT has something like preferred partners or preferred places where it searches um... Then it solves the live data issue and the life pricing issue because to Dick's Sporting Goods database. And you can then confirm that it is in stock and have the right price, which already corrects a big deficit. However In terms of diversity of search results It would be very interesting To see what happens.

**[00:38:44]** So add more to this With chatGPT-Five point three and five point four. That will launch recently Particularly with ChatGPt Five Point Four What I did See Already To be clear, because people on LinkedIn are very quick to attack this angle. ChatGPT five point four has not been rolled out for log dot users. This data can only be asked if you do it in your own personal logged in sessions Which is what I did In my own case where I analyzed effectively My own streaming data. So there's no data privacy issues There at all. Um i effectively recreated asking For different product carousels And I just wanted to see What happened. Now, what I saw and what I've posted about publicly on LinkedIn already is that now with ChatGPT-Fivepointfour, ChatGpT actually has an inherent brand bias.

**[00:39:38]** So it started to... Whereas in my study the shopping query fanouts did not mention a brand typically very rarely so would say best running shoes under five hundred dollars or best smartphone under four hundred dollars something like this if you notice there It doesn't say some specific brand and therefore it doesn't really influence the Google shopping results that are returned. In five point four, this is actually different in five point for there is a brand bias And you can see it in The reasoning steps that it does so when it starts to think? It will often already be thinking of specific Specific brands and this affects the whole kind of like dominoes falling over. it effects the whole chain off.

**[00:40:22]** If it's already thinking that these are the best providers in this field, This affects what products It retrieves. This effects What context it retrieves and with this you can see kind of how sensitive The model is. And I do wonder Is there sort Of super brand bias? That coming into five point four Related to chat GPT Kind of making a move To onboarding clients Uh on their own product feeds. So yeah just food for thought. i don't know the answer to it Already. I think there's a lot to process and a lot of food for thought. A very simple question, so thanks so much for explaining it in so much detail-a simple question around the whole genetic checkout or like integrated check out. also do you thing that OpenAir just identified?

**[00:41:09]** people liked to use chat GPT for product discovery? This is why they have this pro carousel feature etc. But that people are still hesitant to like, basically check out from there. or do you feel maybe it's just too much effort for them? To get a try technically so they had to deparitize. So Just a little bit of your thoughts because You've really been deep into the whole topic. Yeah It is great question! So its'a bit Like The Question For You As A User If You Know That You Shop Regularly On Let'S Say Amazon Or Best Buy If We Take The US Data. So if I go to ChatGPT and drop in the Best Buy search plugin, or however it's integrated now because it does actually change over time. And i know that with this... ...I

get the live product data of the LiveReviewData an effectively more integration than if I ask ChatGPPT to do it natively.

**[00:42:03]** For me then its kind like a no-brainer for me as user If Im aware of the benefit.. ..I would do This! And so for Me The Retailers will always have better Data effectively someone scraping the results. So, for me it's quite... It definitely kind of has to happen that they move more towards retailers because They were not able already in their last one or two years To come up with an in-house solution That says okay we'll aggregate all these diverse feeds Of data. No! They tried this and failed. They scraped Google Shopping and it doesn't really produce necessarily the best user experience if pricing is off or its not technically live data, which is kind of strange. If you're missing out on offers, coupon codes discounts things like this.

**[00:42:53]** And so I do think that makes complete sense from a user perspective That The goal for team must be to basically give better shopping results To users. So yeah This part one of answer Part two Yes, there is data that shows that while a lot of uses use AI for the discovery and research phase. There was still a little bit of lack of trustworthiness on it. maybe users as second step might go to the brand website and check hey or they might read reviews on various social sites in their mind themselves put all of the information together before making purchase decision. they might watch review video on YouTube. so I think having it all in house and one AI window. I don't think there's enough users trusting this end-to-end workflow yet, which is why we saw Agentsic not really basically take off.

**[00:43:48]** Before we go into the practical implications for eCommerce managers because you also put together a great piece on what people can actually Take from this research In order to optimize their presence? i'd just like to ask You if there's anything surprising for you or something where if like, hey this is something that I don't want to be missed. If we talk about the whole research and analysis part before moving over? Yeah sure so yeah i just wanna be so loud and clear that they study does prove the chat. GPT scrapes Google Shopping And for me, this is like the scene in The Wizard of Oz where someone peers behind a curtain and you see that it's not actually magic anymore. It is just scraping! Too often in the case of open AI... ...the answer is just scrapping and often connected with Google.

**[00:44:48]** I can't even call it laziness from the side of engineers because its an incredibly hard task to build a shopping graph. However for companies with such funding from personal LinkedIn messages and congratulations I've received on the study was surprised just how blatant and obvious this scraping was. This is one part, And The other part Is if you are using AI as a user to do your shopping i would say keep A little bit of that untrusting side Of You & Do Go Out & Verify the information, the leaks. Do Your Own Product Research Use AI As A Tool but also actually make sure that you are checking the information is accurate yourself, kind of main points. Love The Wizard Of Oz comparison. I think That's really great visual to have in mind.

**[00:45:41]** So now talking about practical implications because obviously everybody is like okay i thought it was magic. Now know its not. so then lets get to work. how can we with this? What would your recommend? people that work in Ecom, SEO AI search. They want to get their products with so into the product carousels of chativity. where should they start? Yeah So The first part to check I would say is what i would call feed optimization. That simply means for the key product terms or the key searching terms Are you appearing In the top few results of Google shopping? If you want to take it another level, You can then check how high up the results. Your appearing Another level is to check if your only interested about your products showing Up and your product show of across multiple merchants in multiple retailers.

**[00:46:41]** That's another level to check as well. so It really depends on your use case right? So if your use cases we are the brand And We Are The retailer that's quite specific. you need To kind Of go out

and Check firstly on Google Shopping Feed for the locations that you're interested in, for product terms and search terms. Do You Show Up? And that is basically feed hygiene. are your prices up to date as your Merchant Center account up-to-date? Basically all of the basics of Google Shoppings done very well This a sort level one. The other part using platforms like PKI or similar Is also checking If you are appearing already in Google Shopping, does this translate as I predicted it would to carousel positions?

**[00:47:31]** So Are You Regularly Appearing In The Top Five Positions In Google Shoping? if so. You Could Quite Rightly Expect To Appear In The top Few Positions Of the Carousel But You Actually Need To Go Out And Check and Confirm That That's The Case! If It Is The Case Probably Your Doing Quite A Few Things Right Already And There'S Maybe Less For You To Worry About. but you can always optimize further. So actually collecting that data is, something that you need to take the time to do... To basically check it out. In this study we saw a position correlation so if you rank higher in Google Shopping You rank higher on the Judgeability Carousel as well. It's not just let say top-twenty results But ideally Top five would be great.

**[00:48:16]** Your chance of making into the carousel for the related query then will be pretty high Consistency, are you regularly showing up in the top few results of Google Shopping or is it changing over time? Is a very dynamic market that your'e in. So thats kinda where I would start with that and this second part its not just the organic shopping results influencing the carousel but also all supporting information. we could call essentially all of the sources that influence those context queries. So buying guide style content, is your product and brand mentioned there regularly? And if so as it mention more than your competitors in a positive way? with this information you effectively then have roadmap for what to optimize over next weeks or months.

**[00:49:08]** basically For example, so if I'm a shop that sells products that are not necessarily from myself. So I am NOT the manufacturer but maybe just in quotes a retailer and i'm selling for example ultra wide or like high quality monitors? Basically I should make sure my product and Google Merchant Center feed is well optimized Organic shopping results. because I mean, I contributed to the piece you published. This is piece number one and then peace. Number two, I could create content on my own side like a comparison guide duel. so should you go with dual monitors or wasn't within ultra wide monitor? And they should check. how am i mentioned to have these products mentioned across the web? maybe unread it may be YouTube etc.

**[00:50:31]** Is that correct summarization? So You mentioned on your own site would recommend first look at the niche that you're in let's say a high authority, low content volume niche which means something like cyber security. It takes a lot of knowledge to write about how to install a mesh VPN securely on your enterprise software or A lot of the articles are quite long tail and you actually need a cybersecurity expert To even check that it's correct before publishing. In those type of niches we see brands both being The source and the mentioned brand Which is sort of They control the narrative from end-to-end, you know. An example for people to check I'm not affiliated with them in any way but i use it as an example a lot is this cybersecurity brand Sentinel One.

**[00:51:20]** they produce A LOT of that type Of content and You see Them both mentioned As top Cyber Security Brand And The Top Source. However If your In the Industry of Fragrances & Beauty You Know Your Selling Perfume Or Clothes This Is Probably Impossible To Achieve. So...I Would Say That Ninety percent of the content in a sort-of high competitive market is going to come from third party sources. now That means that you want to find a way To influence how your brand and products appear on those sources And that will give you The product and brand mention and sentiment uptick. That then feeds into the AI creating its Bioguide content, and so on. Now there's another level to this that is often missed an overlooked.

**[00:52:04]** So you need to understand the sort of market landscape off those third party sources that are regularly being mentioned. So firstly, I must know who were the top sources? Are they more high quality editorial pieces? or maybe let's say medium-quality listicles to low-quality listsicles like... The fact is whichever sources our ranking are influencing your markets. so it's not really a place to necessarily judge them. It's our place then basically see if we can get mentiond in those Ideally in an organic way. if it's not paid to play, however In the industry of fragrance and beauty you'll find that The percentage Of affiliate content is incredibly high Which actually gives all those brands the lever To pull where If they're Not mentioned They can Actually Then have a Paid strategy too become Mentioned And then will see A direct uptick.

**[00:52:54]** How often they are Mentioning That Retrieval Content which Combines Together In Magic Recipes Who Hopefully Be Featured More the product selection. Now maybe a non-obvious stupid question, can I also get into the Chagabiti Pro carousel by just running Google ads? So you always said it's about the Google organic shopping results so that none paid... Yeah because i was wondering why is Chagapiti not looking at the ads? yeah this. Well, there's a few reasons really but the true reason is it scrapes the organic results because it gives this nice fairly low bias selection of Forty products to choose from. In the study I went to Forty Organic Products Because what i did was i mirrored The exact settings that chatgbt Was using with its scraping provider To get the most accurate results possible.

**[00:53:55]** so That was the top forty organic Results. if you go into google shopping With a US VPN, that will be the part where you see either browse all products or all products completely skipping the ads. Um with ads because they dynamically change so often and the makeup of these ads changes So-so often. if chat gpt just scrapes The ad results it would probably have a worse selection set for the products It wants to feed back into to the carousel. The organic results are more stable over time than, of course dynamic bidding and ads and things like this. And I believe that Prachat Shripti made a decision That this selection of organic products This Forty is better approximation Of what they user may select from Than top few ads in Google Shopping.

**[00:54:57]** Google Shopping is different in different countries, however what's fairly non-controversial is organic products can be shown at most locations whereas ads maybe not. It's not clear that ads could show up on every country so they want to choose something for the majority of locations which are almost always organic. So no way to buy myself into the churchivity carousel by running google shopping ads? Technically, no. Because as far I've seen from now it's not scraping from paid feeds. if you like It would be a better way to Be in the top five-to-ten on the Google shopping feed and have A lot of supporting content. that says If your really trying To rank specific product for a category That says Your product best fits that solution Anything else besides or beyond.

**[00:55:52]** you mentioned that people working in Ecom either for a brand, for retailer selling their own products, selling others' product. That they should think about care about. maybe take from the study if they want to... If they generally want to win-if you wanna call it like this? The AI search presence. Yeah, so one thing is how surprisingly easy it would be to sit an AI agent on top of well-organized data. So in the era of Claude Cote's nanoclaw and running agents relatively harmlessly in docket containers and so on every day It's becoming a little bit easier for people to do that. Their added value is the data modes that they create themselves. And I've been using this example, so you have to bear with me why i choose an abstract example.

**[00:56:57]** but it has a reason. if You think of people shopping for products That are kind-of hard To find and let you need A certain human in the loop Almost like approval to be confident that you're getting It. imagine your looking Luxury Italian furniture. It's a high ticket item, but would you trust chat GPT to deliver your results? Because if it scrapes Google Shopping or scrapes the general internet which is of

course by definition not specific The results aren't gonna be good. However If there's a brand out that who has taken time and energy To collect this information and systemize it, and effectively make their own data modes. If you imagine an agent sitting on top of there helping to make personal decisions likely its going be much better.

**[00:57:42]** so I really believe that i would love to see effectively rather than chatGBT being the default starting place brands begin to win back a bit of their visibility by creating amazing agenetic experiences ontopofthedata in thier own expertise. love to see this year, because if you have well-organized life price data that either already integrated with a Shopify store. You have ninety percent of the recipe that effectively needs your own agent and then it becomes about things like brand awareness. why should they go easier than probably before to get a lot of eyeballs in front of amazing UX experiences that you create. And yeah, just to reiterate the data mode that you created yourself is actually often value and just give it away.

**[00:58:39]** so have an agent check out within AI may not necessarily be best decision for your category. Tom that's a very great summarization, and I really appreciate you sharing such Details about the study now. The obvious maybe not final question but like an obvious question i have to ask is What on the horizon? Like What new research are you working on where? You can maybe give us a little bit of a teaser, A Little Bit Of Your Spoiler so that people... I don't know if PKI already has like a newsletter or something. But yeah! i think you already have a newsletter That People Definitely Subscribe To The Newsletter Or Want to Follow You To Be the First To Read And See what You Guys Are Working On. If They Follow Any Of The Research Team So Thats Me, Malta and Tomek Rotsky On LinkedIn you will get the direct feed to all of the research that we publish, because for now at least it comes through one of these three channels typically on LinkedIn.

**[00:59:38]** So people can also message me and ask me questions about their research which they love to do! And I'd like to answer those questions. but yeah a sneak peek. what's next? i can give you a sneak peak of what im working right now. um...and i find this really interesting. We talked about Google dynamically changing content, and the example that we had in this podcast was it can change the product title or the product description ever so slightly by dynamically rewriting. The reason is because they believe it offers users a closer answer to what their question is asking for. It does the same with meta descriptions – that little description you add on your blog article that Google uses to effectively rank.

**[01:00:22]** it rewrites these very regularly. Why is this interesting? Well, because we know that ChatGPT scrapes these web results. It means by definition it ingests the rewritten description not to be actual description in source code. So what Google chooses to rewrite your meta-description too Is what actually ChatGPP uses to effectively rerank its content. so I decided do a study on this. Um...I won't reveal much but the idea of studying Firstly, this rewritten meta-description idea to see how much of ChachiPT's answer relies purely on this rewritened meta description. My belief is quite a high percent. The complexity to study list is pretty tricky. I will tell you So that... That it effectively what we're doing because its really valuable information Because if we can say there are some high percentage based on quite small snippet We see a very high reliance on how chat GPT actually begins to think about things.

**[01:01:28]** So this is going be one of the next big research pieces that we're working on. Awesome, so people if you are listening and want hear more about the research. Definitely go follow Tom, Tomek and Malte from the PKI team and watch out for what they are doing. And I will also make sure that Tom makes time to come on a podcast again when we publish our next piece. So basically behind-the-scenes is also reasoning in little bit of the nitty gritty details which obviously people can read but you know how sometimes want it to be read. So Tom, thanks so much for taking the time and

sharing this in so much detail. I am very much looking forward every new research piece that's there to come because so far PKI or the PKI Research Team did not disappoint To say the least.

**[01:02:26]** Thank you Nicholas. We're really trying our best to produce super high quality also want to hopefully inspire other research teams, too. Push it to another level make a potentially peer reviewed or at least have the option to look into the methodologies in a lot of detail something like people like Ethan Smith are also pushing for as well. so we really wanna go in this direction As much possible add some transparency To how we do research and the industry as a whole. So yeah thanks Nicholas was a really super fun interview. Thanks so much Tom. Have a great day And see you soon!