096 - AI Content Detection with Jonathan Gillham

Jonathan Gillham 0:00
If there's some question around, how big can it get? Because big just means more content to chew on. But at what point do you run out of all quality human generated content? And if there's no more human quality human generated content to train on, GPT three was trained on an estimated 10% of the Internet came through at once. Well, you know, we all already the internet, there's plenty of internet that we don't want intelligence to be trained on. These places to go to make you dumber.

SEO Leverage Podcast Intro Music 0:35
Welcome to the SEO Leverage Podcast, where we talk about search, marketing, and conversion.

Gert Mellak 0:45
Welcome back to Seoleverage.com. This is episode 96. My name is Gert Mellak, founder of SEO Leverage. And artificial intelligence is in, is everywhere right now, we have been using this for a long time already, but obviously now this is a different game. Chat GPT has really managed to breakthrough, I think, a lot of noise of the early early tools and now we feel like artificial intelligence tools come up every single day and can do more and more stuff of what we used to buy on Fiverr or hire an intern for. So it's really fun times, specifically for SEO, it's interesting to leverage AI to a certain extent, as we might be hearing here for content creation. I'm curious about where this conversation is going to go. I'm very happy to have an expert here in this field, Jonathan Gillham, thank you so much for taking some time for us.

Jonathan Gillham 1:37
Yeah, thanks for having me, happy to, you know, have geeking out on this topic for months now. So happy to happy to be chatting about it.

Gert Mellak 1:44
It's definitely a rabbit hole. So you're the founder of originality.ai, AI detection and plagiarism detection tool. We're going to talk about more about how this might work. I would just like you, to ask you to give a brief overview be, where do you come from? How can can people understand why you have experienced in this space?

Jonathan Gillham 2:05
My background is in web publishing. So I've been building websites for over a decade, had an content marketing agency, had a couple of content marketing agencies that we sold. And so we were sort of very in tune, sort of building building content for our own sites, and strategy for our

own sites and then as well for clients. And yeah, sort of had some pain associated with the plagiarism checkers of sort of, of the day with, with sort of Copyscape being default and wanted to build something that was more robust and useful for a team or sort of hundreds of writers and dozens of editors. And that was, that was what originality was originally gonna be. And then yeah, we sort of saw the need, as well for trying to help educate people on the the source of their content creation. This was sort of pre, pre Chat GPT, but creating a an AI detection tool to allow people to understand the true sort of nature and the true originality of their content, both being not copied, and, and human human generated.

Gert Mellak 3:08
That's amazing. Copyscape definitely has been a tool we have been using for quite a while as well, definitely a default go to tool. But I do understand absolutely, we have also come across tons of limitations. And I'm really curious if they are going to make this step to AI as well, at some point probably they need to do. Before we jump into some, some real applications, what if I read the text today on the internet and this is a question I get from clients all the time, these days. If they read a text on a web page, can they know if this was written by a human being?

Jonathan Gillham 3:39
Yeah, so with with, there's our tool, there's there's others? With the answer, the answer is with a reasonable degree of certainty. Yes, you can run it through an AI detector and get a 94% accurate answer on whether or not that was AI generated. It does have false positives, it does have false negatives, but with some degree of certainty, you can, you can know that . It's AI detection is not as sort of binary as plagiarism checking, where it's, here's your 10 words that are copied, here's where the source was, that is simple. You know, it's like hey, why did this article tagged, get tagged as AI? That becomes a little bit trickier to give a clear direct, these are your 10 words as to why it occurred. So the short answer is yes. Long answer is it is a little bit more nuanced and it's a probability question that at some point.

Gert Mellak 4:39
Right. It was it was interesting before when Chat GPT really kicked in on social media with the first examples and what it can do and what you, what you can done, do now much faster etc. I had a lot of clients tell me it's probably copying stuff together from different websites. But as we understand now, it's really trying apparently to connect the concepts in a large clear way trying to understand to a certain extent how it makes sense. And you can just play with it a little bit, you can copy and paste those texts into Google, you're not going to find them like this anywhere. So obviously, it's it's definitely a different game. Where do you see today, some use cases where AI content makes the most sense?

Jonathan Gillham 5:21
Yeah. So although we have sort of, you know, what it says, at this point, the leading, leading edge detector on the market, I'm not against AI content. I use it on some of my sites I'm not against it. I think there's a huge case for it. I think there's a, you know, certainly if there's only that's definitely being used and being spot checked because it can be a, you know, world class BS or in terms of getting a little bit going a little bit, hallucinate Google is hallucinating on some

different ideas. So in terms of use cases, I think when there's a reasonably easy to spot check, formulaic, there's specifically use case for content marketing, is when it's sort of a reasonably formulaic set of content strategy that you're wanting to put out, that it can go deep on that topic, and able to have a human that can easily value check. So for example, if there's sort of a standard question around what like, say, traits of a certain animal, that would be sort of formulaic enough, easy enough to spot check, and then able to have Chat GPT put some rich content around it. I think that's that's a solid use case. In terms of like, complete creation of the maj, like majority of the content being completely created by Chat GPT, I think in terms of use case, it's an awesome research tool, awesome, brief idea generator, you know, it does it does so many things. So I think in terms of like, bulk creation of content, that sort of a little bit more formulaic, easy to spot check, not mission critical, not your money, your life. Content is where has the most applicability for full article creation.

Gert Mellak 6:57
It's really interesting. We, I think we started probably a year and a half with really taking AI into demand daily processes like content creation, idea generation, title generation, and all those kinds of things. We were very conservative at the beginning, but something we found was that really the research time went down considerably because we were touching on as an SEO agency, obviously, I don't know construction in Southern California, sustainable regulations and stuff like that. You need a writer to really be exposed to this for a few days, probably to start making sense of the different concepts. We're not an industry expert. But then with an AI if you, as long as the prompt is specific enough, like you say, you release, and the download will actually read can't really leave and has enough information to connect the dots and make it very, very specific. The results are astounding, astounding. And one of the main reuses cases I think we found apart from title generation, or edit text is seems to be working really well, was really just outlining like an article or a marketing strategy or like a book outline for XYZ, etc. Just to really generate those ideas, it seems to be really interesting.

Jonathan Gillham 8:01
Yeah, that like top goal in the world of SEO or a top website maps and topical authority is a very hot topic right now. And I think it's, you know, that like topical authority is a sort of cascading thing that flows all the way down to how is this article the sort of topical authority on this specific longtail, longtail phrase, and you know, there's been tools for a while, like the surfers Americans uses that sort of help guide to that. But I find this sort of the brief creation out of Chat GPT gives you a more structured approach to sort of getting that completely covering a topic and achieving that sort of topical authority, but with a brief from Chat GPT, then from some of the tools that we've used historically for that. And I think that's that sort of like, it's applying that intel, you know, so applying a level of intelligence to it, that we don't get from sort of the road here. You have the keywords that exist on other articles, puked onto a page.

Gert Mellak 8:53
Absolutely. Well, what we found really funny, we played with this a little bit was actually adding a de-existing text from very experienced writers. So I would have my senior writer, write a paragraph, and then after GPT to rewrite it. And you will see how we extract concepts, right?You

could almost see the process and the list, suddenly, we have shorter sentences, we have easier structures, we don't have commas pretty much because it's, they really, really go like directly for the relations, and XYZ is ABC, and stuff like that. So it's like very simple structure as like, this is apparently the kind of text we should be writing, right? It's not only easy for the machine to understand, but it's also very real to read. So it seems like there's only there's mostly upsides in this, in this regard. We are very conservative as an SEO agency, so we always give like the advice to have really a human element, make sure the brand voice is there to edit. It's user friendly, there's some empathy, knowledgeable etc. So I guess the emotional element is hopefully the last one they're going to to conquer. It's really funny how this works. How does the detection actually work? Can you explain this somehow so non tech experts would, would understand the process?

Jonathan Gillham 9:59
Yeah, so the way we built it was we saw an open AI model that they make available, we trained it on a crap ton of output from GPT3 and GPT3.5. And then sort of trained it on both human's dataset and a AI data set and said, you know, AI or AI, so like originality that AI is AI learn what this looks like, and the differences from this sort of there's human verified human to verified AI. And it's a sort of the analysis of like, you know, the question often comes up, like, Hey, why is this article AI? And, you know, the answer is like, we don't, we don't like because they as I said, we've tested the accuracy of it, but we don't, we can't say with certainty as to like, hey, these were the 10 words that triggered that and say sort of a similar analogy is like with sort of like trading firms where they have bots that are deciding on and executing trades. A human wouldn't necessarily be able to get in there and say, like, okay, it's within our risk parameters, but I can't say exactly why that trade is the one that's getting triggered. And it's similar within sort of our AIs, we will get, our AI has been trained has been sort of verified to that level of accuracy. And it's, has caught things within this. So one thing that's different about our tool compared to some of the GPT2 detectors, is they look at previous detectors look at like, hey, what's the probability that the next word was the word that AI would have said to be, and it becomes a very sort of like, linear probabilistic problem. Whereas our AI model will look at an entire article more holistically, run an entire set of texts more holistically, and say, okay, because of these seemingly, or to humanize seemingly unrelated parts of the article, we've determined it to be AI.

Gert Mellak 11:51
Fascinating, really fascinating. So it's, it seems like the more you train your than your model on more GPT3 content or GPT4 maybe, the more precise it gets, this is how this works?

Jonathan Gillham 12:04
Correct? Yeah, we started we trained it up to the point of seeing sort of diminishing returns on on level of, level of accuracy, but yeah, like, what will will we need to feed it 100 times more data for GPT4 when it comes out, I don't know, we're gonna, we're gonna find out.

Gert Mellak 12:20

Absolutely. I guess with your background in content marketing, publishing, etc, I would really be interested, in your opinion, what's going to happen in the next six to 12 months? With Google, with content interpretation, we can't just guess at this point, right? I've got my opinion, I really appreciate you sharing yours. Maybe in six to 12 months, we were still in touch and, and either are surprised by the accuracy of our estimations or really, really surprised by what Google is doing. What do you think, how are they going to react to content being a commodity to everybody, being able to create a website about cancer treatment and surgery and financial advice?

Jonathan Gillham 12:55
I think they have to respond. I think I think it's an existential threat. I think they've been very careful to say like, Hey, we, you know, like, there, there are sort of sequence of information that they have shared this, maybe Google has said, hey, just a reminder, spam generated content is bad in whatever form it comes in. And then by the way, we are capable of detecting AI content, and then coming out and saying, like, just so we're clear reminder of point number one, we have never said that AI content is bad. You said that spam generated, no value added content is bad. And so I think what that means to me, and I think so like if you think a lot and you think about like, What's Google trying to do, you know, ultimately make money. And I do think they are facing for the first time in a decade and existential threat. Where I'm going to charge GPT for some searches, as I think are others where I would have historically gone to Google. And so that's going to start taking a bit of market share, not, you know, probably meaningless at this point. But then the other problem that they're facing on the other side is if their search results get absolutely flooded with AI generated content, will they need to apply some mechanism to try and keep their search results as a human generated? Like if everything is, if everything that is, you know, I think no, if you would ask anyone, right, like, you're not talking to Chat GPT for this podcast, you're talking to a human. I think people want information from experts. They're okay with it being AI assisted information. But you know, I think the focus on E, if you're gonna be coming, so if I were to make like specific predictions, that focus on E E A T is going to become even more significant. In 2023. I think they are going to continue to suppress AI generated content, I think that's going to have to happen for them. And I think they might launch their own version of check GPT if I were to go out on a limb, those would be my kind of predictions over the next, because they have the biggest in terms of number of parameters. They have the biggest NLP model in the world.

Gert Mellak 14:52
That is fascinating. I think it was it was surprising they wouldn't react any earlier or bring something out like it seems like they want to do to pull the brakes on it for a while. because we have seen in SEO, they have to have been working towards this for 10 years, or longer, where they just focus even more on the author and, and the E E A T scheme, the E E A T scheme now. It's so funny that it seems like links are going to be the only way to figure out if something is legit at some point. So they have been trying to get rid of links. Now the best thing you can do is get a link from an authority that's trusted to get your your content verified, and backed up essentially. It's really fascinating. So we what we think here are what I believe might be happening is that it definitely just add it like into the Search Console, issues and warnings

ahead seem like you heavily rely on AI content here or they elect to do with con with covert vitals basically penalizing you just because your site is loaded, just start issuing some warnings here and there say, Okay, this is now a ranking factor. We look at this, and what we think about this article and these kinds of things, maybe to, to contract a little bit, but obviously the, I mean, there's still one, one main drawback I see from Chat GPT. And this is like the lack of source. But I mean, there are also still already solutions out there that start putting featured snippets together with the references where information comes from etc. So it's definitely something that can be solved. And once you can really trust this information, because it's contrasted, I would believe Google needs to have an alternative solution. Yeah. Yeah.

Jonathan Gillham 16:22
Yeah. I hadn't thought of the sort of the applying the Google like the web console, like the search vitals in a similar search experience. Score is too similar to being like a content, I haven't content, a content score.

Gert Mellak 16:36
It would be it would be following the normal structures, as I believe. You'd read it was always like, now we have a mobile focus, we issue mobile warnings, if you're not complying, we have a speed focus, we issue covered vitals warnings, right? We have a link spam algorithm, we issue manual penalties or warnings or, or stuff like that. So it seems like the next logical step would be starting there, even if it's not really intrusive. And this is another ranking factor, which is start detecting this. So you know, we are behind this, etc, to maybe pull the brakes a little bit on too much. But it's generation. But I agree, I believe they have to act, they have to act in the best interest of everybody. If you don't can if you can't trust any information anymore. It's really, really tricky. And definitely competitors are embracing this, like crazy deflectors finally, this disadvantage they could have timing wise, etc. And I think everybody everybody loves Google. Okay, so many people love Google. But I'm okay with a different tool , as well. Like I said,

Jonathan Gillham 17:33
I agree. I was like, like, what are the moves that Microsoft is making them like and what like, you know, if you were to go back 20 years, and what, what world are we rooting for Microsoft sort of the underdog? Like, it's like, wait, Microsoft's the the underdog against Google? They go back 20 years? And that'd be pretty funny statement, 30 Year, yeah.

Gert Mellak 17:58
100% I'm still waiting for the for the voice assistants or home assistants to jump on this a little bit more. That seems like they're taking too long, I would expect the Lexa to already be be able to give much better answers, but in by this time, but they're also apparently we're waiting on on where this is going to go. The one thing that's in the media, I love to see if you can if you can speak to this a little bit in a few months. GPT4 versus GPT3. Can you give us an overview of what's to be expected from GPT4? It should be like the best thing ever, compared to version three, how good can it get?

Jonathan Gillham 18:32

Sam Altman, the CEO of Open AI maybe we've been chatting quite closely because it's going to impact us significantly. There's some sort of well popularly shared images showing the number of parameters compared to GBT3 and you know, he said that that's just total total BS, no idea where that came from, they haven't shared that information. So if you were to look at the number, the sort of the lead and size between GPT2, GPT3 to then potentially where GPT4 would go, will it be 100 to one larger in terms of number of parameters? There's some question around, how big can it get? Because big just means more content to chew on. But at what point do you run out of all quality, human generated content? And if there's no more human quality, human generated content to train on GPT three was trained on an estimated 10% of the Internet came through at once? Well, you know, we all know the internet is there's plenty of internet that we don't want intelligence to be trained on places to go to make you dumber. And so it can't train 100 to one more if it's already been trained on 10%. So, you know, they've come out and said, like, you know, there's a lot of people that are GPT3's, that Chat GPT is so exciting right now that there's people that are saying like, Hey, here's what the future will be. And basically for all of the massive excitement that around it, prepare to be disappointed. Not that it might be hands-on or under promising to over deliver, but could it be incredible like If you look at the content that got created by GPT2, it was garbage, right? This conversation would not be happening, it was clear crap. GPT3 change the game, Chat GPT but an intelligence around the structuring of that data that greatly outstripped the quality of just grammatically correct written words that came with GPT3, so well, we'll be more of that and will then be indistinguishable from human. I don't know, I don't think it will. I think any of the extreme expectations is not going to hit those, it's not going to have the extreme expectations that people have of it. But I think it's going to be increase, improved level of performance on GPT3, of course, and is going to be on the same order of magnitude improvement of going from GPT 2 to GPT3 and if so, that's going to be a huge game changer. In terms of ability to detect it GPT 2 detectors still work reasonably well on GPT3 generated content, so that sort of has me a little bit confident that we'll be able to still detect it. But one of the graphs that it'll be hard to communicate via audio, but one of the graphs that had me, had me , has been the most you're thinking like, holy crap, this could be different is, if you look at like, a curve showing like, quality of content, and the band that exists for free, you know, we think of the best content as the what the best human can generate. But that's not necessarily the case that right, it's an, how good can it get is a little bit more unlimited than that. And so we look at sort of a graph that says, like, monkey hitting keys, okay, now look at like kindergartener hitting keys and maybe saying, Hi, mom. And then you get to like, that band. There's that band of like, what is human interpretable or human generated? Both quality of content is very, very narrow. There's a lot of crap below it of like, you know, monkeys hitting keyboards, but then how high above that band of the best writer do we go for for AI? And that's sort of like the 'Prepare to be disappointed' part, I think, because it's limited right now on training based on human generated content. And there's only so much of it, and there's only so much good stuff and Chat GPT3, or GBT3 already chewed up 10% of it. So yeah, that's, that's, I would say my, my take on GPT4, my long long take on GPT4.

Gert Mellak 22:27
That's very interesting, yeah. That's very interesting. I guess there's, there's probably also, I mean, now, now that we're talking about starting to charge for, for AI, we've connected our own

SEO platform to connecting it to Open AI, obviously, then charging for credits and charging for processing time. So I think people slowly are going to realize there is some value they need to compensate, as well. And they would expect there's also some sort of diminishing return on, on a platform like Open AI to actually do even more processing, how, how good can it actually get? And when we when we take out what what Google very often seems to be doing is just really show, I mean, what is good content, right? What is good content, apart from certain parameters you can take? But what is good content? It really depends on what do they, those who Google Search users achieve right now. We've just had a really crappy page showing number one for the height of the Eiffel Tower, that didn't touch this page for 20 to 20 years, probably, it's still the best result, because Google just knows exactly what the user is happy with. And they really don't care about anything else, which might open up the possibility for Google maybe to accept certain AI content, if users are happy with it. Really, ultimately, let the user have a say with their signals and their behavior on, on what is actually deemed as, as good content here.

Jonathan Gillham 23:45
Yeah, it's going to be me interesting. I'm excited for it.

Gert Mellak 23:48
Absolutely going to be interesting. Before we wrap this up, I would like to ask you to maybe briefly explain how your software works. So if I have a piece of content, if I have a, I want to get a piece of content from my writer, from my agency, from whoever. How can I use your software in order to get an idea if this was AI written or not? And what's going to be output? What is it going to tell me?

Jonathan Gillham 24:12
Yeah, so it works. It works. If you know your audience be familiar with like plagiarism checking tools, it works very similar to that it's combined plagiarism checker and AI detector. And so you can use either the copy and paste function that you'll be familiar with from plagiarism checking tools, or we also have an API. And so you're putting your content hit scan. And then we have two modules, the plagiarism checker on the AI detector, both go and do their work, and then come back with a score, identifying if it's suspected of being AI, AI generated, and then in terms of how would people use that? It's really about sort of getting that sort of information, disadvantaged information asymmetry between like, you got a writer, you got an agency and you got a client. And I think a lot of people are happy to be paying, happy to be paying for human generated quality content and also have to pay for AI cont, generated content. But there's a difference in the audio and sort of making sure that people are people are aware of what they're what they're purchasing, whether that be an agency buying it from a writer, whether it being a web publisher, buying it from a writer, whether they're being a, a, a, again, a web publisher, but buying it from an agency. Sort of trying to drive that sort of knowledge of as a human generated or not human generated, and then just paying it, paying through that fair, the fair price for the for the value to not be so knowing getting an disproportionate amount of amount of knowledge. And like I said, I think there's a place for both. But everyone's better off knowing the true source of that of whether that kind of was created. So pasting the content hit scan and gives you a result very similar to a plagiarism checker.

Gert Mellak 25:47
So this is like a probability this is at a deeper, is it likely to be to be AI content, or half of this AI content half is not or how this is working.

Jonathan Gillham 25:56
Yeah, for the probability. So they'll say something like 75% AI. So that's 70% confident that this content was AI generated, and then it won't yet so we're adding a feature to highlight what parts of the contents or what blocks of text had would have had the highest individual scores to help sort of identify, Okay, this was probably the portion that that we think was was AI generated.

Gert Mellak 26:12
Beautiful! I was thinking about reading that right now, if somebody takes an AI content, or has the individual paragraphs of the text created by AI to save some research and other paragraphs, I'm not AI created this would just factor in right now in the overall probability, I guess, right?

Jonathan Gillham 26:38
Correct. Yeah. So yeah, trying to try to be able to drive that down to the to that sort of block, the block level. And then I think one of the, like, if it's 90%, human 10% AI, then that's, you know, for, and we've got to do a little bit better communicating what those scores mean. But those scores really show, okay, this, like, if it's 90%, human 10% AI, really, really good chance that that was that was human generated, you're 90% chance, especially if you, it's best if you can sort of look across a large sampling of content from the same writer to get a sense on it. If they're always hitting 75% AI, probably some edited AI content being used. If it's always like 90,90, 90 human with like one hitting 50% AI, like 50,50, where the tool is like, I don't know, but the rest of the time it's saying 90% confident this was human generated. Pretty good chance that writer is not using, not using AI to generate its content.

Gert Mellak 27:32
Awesome. Perfect. Perfect very, very nice, this is originality.ai. We are going to link to this in the show notes on and the written summary on SEO leverage.com forward slash podcast, we have episode 96 with Jonathan Gilliam here. Thank you so much, Jonathan. You've given us a pretty good masterclass, I think around AI content creation, detection and stuff like that. So there's definitely spot on for what people are hearing all the time reading about all the time. If somebody wants to get in touch with you to see how they can work with you, your tool, what's the best place to do this?

Jonathan Gillham 28:05
Yeah, best place is that originality.ai Yeah, they're sort of helping, helping the team move the project along as fast as we possibly can. So yeah, any any email at originality, I will eventually eventually get to me if it's requested or or you can find me on LinkedIn as well.

Gert Mellak 28:22

Perfect. Amazing. Thank you so much for taking the time walking us through this. We're going to publish this soon to get this out timely. And everybody encouraged you take out originality.ai, I will definitely see some random tests on our own writers' work and see what the tool gives us. And we all excited to see where AI is going to go. Thank you so much, John.

Jonathan Gillham 28:43
Likewise, yeah, thank you. I was always fun to talk on this topic.