Cody Madison Podcast Transcription
CS0202073
Total Audio Length: 00:63:27

Fowler: [00:00:04] Welcome to Perpetual Motion, an Innovation and Invention podcast. I am Colin Fowler.

Glenn: [00:00:10] And I'm Michael Glenn. We're patent attorneys at Perkins Coie, and we're your hosts for this podcast.

Fowler: [00:00:16] On this show, we discuss technology that intrigues and fascinates us, along with issues relating to inventors.

Glenn: [00:00:22] While we are attorneys who endeavor to gain protection for our client's inventions, we're here to connect with interesting people who have interesting ideas.

Fowler: [00:00:29] Because we are lawyers, we do have to note that nothing stated here constitutes legal advice and that the views expressed here are our own and not the views of our firm, Perkins Coie. Hello. I'm Colin Fowler.

Glenn: [00:00:44] And I'm Michael Glenn.

Fowler: [00:00:45] Today we're doing a special episode with the USPTO, that is the United States Patent and Trademark Office. We have Jerry Ma, the Director of Emerging Technology, with us today. Jerry, can you tell us a little bit about what you do at the USPTO?

Ma: [00:00:58] Colin, thanks so much for welcoming me on to your show. I'd love to give an introduction both as to my role as to the broader ambit of what we're trying to do at the USPTO in relation to emerging technology. So, I'm Jerry Ma from the USPTO, where I serve as the Director of Emerging Technology. And as many of our listeners know, the USPTO is the agency in the US government charged with administering the nation's patent and trademark systems. So, through our work, we aim to foster scientific progress and economic growth through the provision of reliable, timely and high-quality intellectual property rights. And naturally, artificial intelligence, machine learning and other new and emerging technologies are becoming an increasingly consequential piece of that mission. So, my job at the agency is to oversee a portfolio of initiatives and efforts in the AI machine learning and emerging technology space toward ultimately better achieving our agency's mission and our statutory and constitutional mandate to promote scientific progress, innovation and economic growth in this country.

Glenn: [00:02:15] Excellent. Can you tell us about the Patent Office's early efforts in emerging technology?

Ma: [00:02:21] Yeah, I'd love to. I'll get started by describing our overall portfolio of work in AI and emerging technology. And I'll do so by noting at the outset that we're really one of the few federal agencies in which AI and emerging technology is just such a whole of agency effort spanning all of our business units and of course heavily prioritized by IP, you know, the highest ranks of agency leadership. And I'll start by describing the three principal roles through which the USPTO interacts with the AI and emerging technology ecosystem. You know, broadly speaking, we wear the following hats in this world. The first hat is that of an IP office. The second is that of a consumer of AI and emerging technology. And the third is as a contributor toward AI and emerging technology. Now, if you think about the first role I just mentioned that of an IP office, that role is perhaps the most self-evident, at least for those in the intellectual property space. Our agency, you know, through the faithful and administration of the intellectual property system, we provide incentives for the investment and research, development and commercialization that's ultimately necessary to maintain our nation's competitiveness in AI and emerging technology. And to that end, our agency leadership has really convened experts from across the agency, all of the USPTO business units, to figure out the policies, research needs and other initiatives to really make sure we're doing all we can to ensure we're performing this role well, and that the resulting innovation we see out in the public has the best chances of solving crucial world problems. But beyond our role as an IP office, we also look toward the AI and emerging technology community to better achieve our mission as an agency. Specifically, we really, we really seek to harness the promise of AI and emerging technology to better serve our stakeholders in the community, which is, again, a whole of agency effort in which we combine the technology expertise that our agency is known for with the intellectual property expertise that we naturally have by virtue of being an IP office to really ensure our efforts in this space best address the needs of our stakeholders and staff, who rely on a bevy of technological tools to simply be able to do their day-to-day work. And lastly, and this is perhaps most surprising to those outside the agency, the USPTO has recently started becoming a contributor to the state of the art, to the research frontier in AI machine learning and other emerging technologies. Specifically, our agency disseminates, and in many cases, we're actually required by law to disseminate, comprehensive information on the history and dynamics of scientific, technical and economic progress in the United States. And this information spanning our patent, both our patent and trademark archives, collectively comprises one of the largest curated datasets for AI and natural language processing research, which has been used by the private sector and by academia to great success. Now, all three of the roles I described are absolutely essential, and I'm really looking forward to exploring them in detail with you over the course of our conversation. I think just, you know, to give a teaser, our early efforts in broad brushstrokes have been a mix of internal and external explorations. So internally we're, you know, we're assessing what the needs of the AI and emerging technology community are, what the ideal role of intellectual property is in that community, and how our current regime of intellectual property influences the dynamics of AI and emerging technology

innovation. As you all are familiar with, our favorite three-digit numbers, 101, 102, 103, 112. Often, I, you know, often come up in the context of how our current intellectual property system laws, regulations can influence either positively or negatively innovation in this field and, you know, what we should do to make sure intellectual property best fosters the progress that we want to see. And we're also engaged in many early-stage technology development activities, some of which have actually been deployed either to the examining court within the office or actually to the public at large to leverage the power of AI and emerging technology in areas such as prior art retrieval and classification of patents and trademarks. And finally, we continuously explore how we can best unleash our agency's resources, such as our open data archives, to actually serve as the basis for the next groundbreaking developments in AI in emerging technology. And I think it's really exciting for me to be able to share that. You know, as of the first week of May, this May, May 2022, the world's largest open source, publicly available AI language model is actually one that's been trained on the USPTO open data archives. So, in short, it's such an exciting time to be at the agency and to be able to see all the progress that we're pushing forward on the policy, technology and research fronts.

Fowler: [00:08:14] So I'm curious about this model you just said it was trained as of, I don't know, yesterday. I'm curious specifically on what that natural language model is being used to interpret. And do you mean by it was trained on just all of the patents that you have publicly available?

Ma: [00:08:33] Sure thing. I'd love to dive a bit deeper into the at least some of the broad, you know, surface area of what we, what we're referring to when we talk about language models and specifically why language modeling is such an important part of AI research and yes, some of the groundbreaking results and headlines that we see as to AI and machine learning. In short AI language modeling is really about creating a model that can conditionally generate language that would be akin to that which a human either says or writes. So, you know, you can think of now starting off a sentence with the cat. Okay, so the cat and then, you know, that's, that by itself as an article and a noun. And then you give that sentence fragment to an AI language model. You know, you just give an AI language model the cat. And the goal of language modeling is to be able to have an AI tell you probabilistically what the next word will be. So, this is called generative language modeling. And the result of running an AI model on, you know, quote unquote, the cat might be, okay, with a probability 40% the next word should be chased, with the probability 20% the next word should be ran, with probability 10% the next word should be napped. You know, so, the cat chased, the cat ran, the cat napped. And, you know, in isolation, if you think about this task, it might not seem all that interesting. But actually, if you, if you just extrapolate many steps of this process that I've just described to you of, you know, filling in the next word in a sentence, in a paragraph, in a longer form piece of prose, you can actually extract a whole ton of meaningful knowledge from an AI model and therefore from the dataset on which the AI model was trained, just by continually prompting the AI model for the next word, for the next sentence,

for the next paragraph, so on and so forth. So AI language modeling has, I want to say, at least since 2020, and to be honest, even before, you know, been one of the predominant research areas in AI due to a confluence of innovations in AI model architecture, that is the internal design of the neural networks underlying the AI models, along with a, you know, an explosion of compute power and most importantly, of large datasets that are suitable for training these language models on. And one of the largest and most high-quality datasets, because everything that goes to the USPTO is ultimately reviewed by human experts is the USPTO's open data archives across both patents and trademarks. So, in short, you know, this AI language model was trained on not exclusively but significantly on the USPTO patent archives. So, you know, the whole history of many, many decades of innovation in this country as summarized in natural language form in US patents that has made its way now to the knowledge base used by the largest publicly available AI language model.

Fowler: [00:12:14] I mean, if you're going to be using that database, I would presume that the, the highest prediction would be the cat includes various embodiments.

Ma: [00:12:21] Exactly. Or, you know, or we claim, you know, we claim a method, we claim a system. We claim, you know, it does, it does many permutations of you know we claim X but, but, fortunately when you're, when you're doing, you know, any sort of language modeling, you're not limited to one exclusive domain. And in fact, the best, most general-purpose language models tend to include data from a wide variety of different domains. So, you know, some, some parts of this dataset that I'm referring to will be patent data. Some other parts will be, you know, maybe fiction books and some other part will be, you know, maybe the English Wikipedia archives and so on and so forth. So, so it's really about trying to get a diversity of generally useful, high-quality knowledge in these sorts of models.

Glenn: [00:13:08] So, so what is an application of this database then. For example, I mean, would an examiner then use the database in the course of doing a search? Maybe he would have a query and then the AI would help complete the query.

Ma: [00:13:22] Yeah. So, the uses of language modeling are manyfold both in the general sort of AI industrial complex, but also specifically in the intellectual property context. And I think one really interesting trend that's emerged from language modeling but really, you know, the, the AI research community in general over the past half decade has been this idea of representation learning, which I'll explain a second. So, representation learning is this idea of note by training AI models to do this sort of like fill in the next word, fill in the missing word sort of task, which intuitively, you know, doesn't seem like it would encode all that much knowledge. It still turns out that just by in, just by training these sorts of models to do these simple, you know, next word prediction tasks, the resulting intermediate representations of language, of words and sentences and so on and so forth in the models, you know, in what we call the hidden layers of the neural networks actually end up being these very rich, semantically meaningful

mathematical vectors and some, you know, ungodly high dimensional mathematical vector space. So oftentimes what you'll see is in these language models, you know, words will get projected, or sentences will get projected into like a 2048 dimension. And yeah, they're all, they're all often powers of two because powers of two are nicer to work with, but, you know, 1024, 2048 are common, you know, dimensionalities for these hidden layers in these neural networks. [00:15:00] And you might think of, you know, what happens by virtue of training the AI language model, you know, dimension one might end up being, you know, is it a noun or not a noun? Dimension two might be is it a verb, not a verb, and then so on and so forth. And as you get more and more dimensions you are able to encode much more nuanced and semantic knowledge as to, you know, what's actually going on in the sentence. What does this sentence actually mean? Dimension 500 might be, you know, is this related to intellectual property at all? And, you know, if you think about using those what we call representations or embeddings, these vectors, you can sort of think of reason about them almost geometrically. So, it might stand to reason that patterns that have similar representation, embeddings or vectors in this high dimensional space are actually, you know, good sources of prior art for each other. So, you can do some sort of like nearest neighbor search in the mathematical metric space in this vector space to, you know, to get a sense of, okay, here's, here's point X, which represents the patent that I, the examiner am trying to examine and search and so on and so forth. And I'm going to get the ten patents that have the closest vector representations to point X, and then I'm going to use those as maybe the basis for at least my initial prior art search. And, you know, digging into potential novelty or obviousness issues based on, you know, just based on pure mathematically, you know, perhaps mathematically complex but still semantically meaningful representations.

Glenn: [00:16:38] Yeah, but unlike keyword searching, which is often brings up things that are unrelated to each other, you're able to be more precise here because you're doing the mathematical comparison. Is that correct? So, when I put a keyword in, in this case, it knows what's going to be relevant and it doesn't bring in things that are like maybe synonyms that are not synonyms, they just simply sound the same, but they're not the same.

Ma: [00:17:00] Exactly, a synonym disambiguation. And you know, I guess the technical term we might use for it is, you know, some, probably something like semantic matching or semantic similarity. That is a task that is super difficult in the general, you know, corpus of English language, but particularly difficult in the patent space because as, as you both know, know the applicant gets to be their own lexicographer and times they get to define terms and words as they wish. And therefore, the variance in definitions of even ostensibly identical or similar terms can be high, bordering on unbounded. So, this task that you're referring to, you know, synonym disambiguation and/or, you know, semantic similarity is actually one of the I think I probably would go so far as to say, you know, one of the holy grails of AI modeling in the patent space.

Glenn: [00:17:57] I think then they wind up getting much better search results. I think it's sometimes frustrating for practitioners and this is not a criticism of the office, but that an examiner will, will find a certain word in the claim and will find lots of prior art that has that same word but in a different context. And this would actually avoid that. And when the examiner was to do a search now based on what's we claiming, this technique would actually make the search results much more meaningful.

Ma: [00:18:21] Yeah, there's a good chance that, you know, what you just described can be facilitated by AI models simply because, you know, under, under the theory that I guess I'm thinking of AI models would tend to encode the language of the claim in a context aware manner that actually, you know, gets a, you know, more solid idea of claim scope than perhaps just a keyword search derived from specific limitations in the claim. And, you know, under that theory, you know, it really would become a more meaningful exercise of, you know, finding prior art that actually relates to the underlying claim scope rather than simply, you know, trying to do a best approximation of that with various words scraped from the limitations therein.

Glenn: [00:19:08] And as the search gets more, more accurate, of course, it's harder for us PAT lawyers to, to avoid the prior art because it's the results are becoming more pinpointed, I think, you know. And so, it's actually, it's better, I think, all around because we're not spending a lot of time on prior art that's not so meaningful. And we're getting right down to the issues more quickly.

Ma: [00:19:26] Yeah, ultimately compact prosecution, I think is, you know, something that we very much try to practice at the office. You know, getting to the crux of the issues, including novelty and obviousness issues in any given application sooner rather than later. And I think, you know, (1) it's as you alluded to, I think better for, you know, all the stakeholders in the ecosystem to get to the meat of the issues as quickly as possible and (2) something, I think AI and emerging technology is especially equipped and suited to help us do at the office.

Fowler: [00:19:59] Do you ever see the sort of modeling to replace existing examination procedures or be in tandem with it?

Ma: [00:20:06] That's a fantastic question and something that comes up, you know, more frequently than I would have expected heading into this role and heading into the PTO. This question of, you know, are we, what are we trying to replace examination or, you know, what specific aspects of examination we're trying to replace? And the answer, no, just bottom line is none of it. So, we very much believe, at least at the USPTO, in making sure that, you know, technology at its best really empowers experts rather than attempt to, you know, supplant them or replace them. And as to AI specifically, I think we at the PTO are now very much pursuing a non-dispositive human first agenda for our own AI activities. This is, you know, due to many reasons. I, you know, one, we really do believe that our examiners and the other experts I know who I'm honored to call my colleagues,

that they are really now the source of expertise as to any substantive, you know, adjudication decision. You know, they're best able to marry, you know, their knowledge of the law with their knowledge of the facts as to any particular application to synthesize that into reason decisions that ultimately advance prosecution forward. So, so, I think, I mean, now maybe one some, one substantive way to ground or to maybe illustrate my point here is, you know, by describing something that you all are probably very familiar with, a noncanonical test for the patentability of a purported invention, which courts have dubbed the Alice/Mayo test, right. So, what, what do you all do? And what do we do under Alice/Mayo while the patent examiner is required to number one determine whether the invention described in the application concerns, I think it's like an abstract idea, a law of nature or a natural phenomenon. And then if that's the case, then do something like determine whether the invention contributes enough, you know, enough beyond the mere idea, law of nature or phenomenon to constitute something, quote unquote, significantly more. And I think, you know, you all are, of course, experts in this space, but one doesn't really need to be particularly, you know, proficient and patent law to intuit that this test often becomes like an incredibly nuanced judgment call. Right? What precisely is an abstract idea? What's a law of nature? What's a natural phenomenon? What does this like significantly more thing really mean? I mean, our courts are really still hard at work drawing the contours of, you know, what these things all mean and made, you know, just exponentially increasing, exponentially exploding complexity in the core science and technology. So, I think the notion that any AI system could correctly make these, you know, adjudications under even one single test, right, even under Alice/Mayo alone, let alone the numerous other determinations that examiners have to make on a day-to-day basis is a bit fantastical, right? So, you might be able to do something like, you know, just make a data set of patent applications and train an AI model to make predictions as to, you know, good under Alice/Mayo, bad under Alice/Mayo. And you might be able to do this from a statistical perspective, and you might even see that you get some good, you know, headline accuracy figure like 70%, 80%. But we're not really in purely the statistical prediction game at the PTO because our job is not only to, you know, make or it's not even at all really to make the best guess as to what should happen. It's really to reason from first principles as to what should happen with any given application based on the law and the facts. And the examiners are reasoning thereupon. And, you know, just, just playing a statistical prediction game as with, you know, many things and what we call supervised machine learning really doesn't satisfy that, you know, really doesn't meet that mandate at all if we're just, you know, given these opaque black box, yes/no determinations even for, not even for the examination as a whole, even for individual facets of examination, then we're really, you know, removing from the examination process any traces of no accountability. You know, learning at its core examination is supposed to be a collaborative process, I think, between the examiner and the applicant or practitioner, where we really try to get to the heart of the issues as quickly as we can and figure out how, you know, we resolve any issues that stand in the way of patent allowance as expeditiously as possible. And

if we're just spitting out these unaccountable, you know, AI, AI generated adjudication decisions, then, you know, that's at the outset there are, there are a million different issues, but that, with that paradigm perhaps but at the outset the foremost issue I can think of is just it's not accountable. And, you know, there's not really anything that either the examiner or the practitioner can do based on just some black box statistical determination like that.

Glenn: [00:25:29] Well, there's also some concern now about AIs learning the prejudices of the people they learn from. And there, there's a lot of attempt, I think, being done in the AI space to try to teach AIs not to become prejudice based on what they're learning from people. I mean, because people are not necessarily ideal in the way they talk. And if the AI gets exposed to a certain amount of human interaction, they may actually pick up the quite unknown or unexpressed prejudices or biases of people. Maybe you could speak to that, too.

Ma: [00:26:02] I think, Michael, you're going to see that issue come up across AI deployments in general, but especially in the federal space over the next decade. So that's an excellent topic to discuss simply because it's gaining more attention from governments, from the public, from civil society. And federal agencies, including the USPTO are very much on point to ensure that any uses of AI and public administration do not contribute to some of those, some of those really scary problems that you mentioned and associated risks. So, I think one good place to start as far as, you know, formal mechanisms the government is using is an Executive Order numbered 13,960. And I forget the exact title, but my best paraphrasing of the title is something like promoting the trustworthy use of AI in the federal government. And it outlines a set of, you know, broad principles that agencies are directed to consider in any AI deployment. And they include, you know, accountability, fairness, transparency, performance, auditability, so on and so forth. And we at the PTO are very much committed to making sure that we really take to heart those principles and any deployments that we, you know, even think of. So, I think we at the PTO are in a bit more of a fortunate position than, you know, most who would seek to use AI in a public sector context simply because our data while, you know, while I don't claim to, I don't make any claims of our data being perfect, I do think our data is comparatively among the more well curated and less prone to biased datasets out there. So, there are some methods of data collection, such as scraping the, you know, scraping the open internet that very readily lead to, you know, bias, hate speech, other things you don't want an AI model to pick up on. And I think we've seen with these general purpose language models trained on things like open internet scrapes, you know, it's been openly acknowledged even by the creators of such models that, hey, there are all these risks when we train our model on the open internet that, you know, some with some non-zero probability that what you're going to get when you feed in, you know, a prompt into the model is something that might have some questionable content, whether it be, you know, hate speech, other, you know, offensive language or even language that just tends to, you know, make unfair construals or disparagements among one class of people or the other that, you know, we really don't want to see from AI models. We fortunately, you

know, are insulated from a lot of that at the PTO. And that isn't to say that, you know, we, we get to just sort of turn our brains off when it comes to dataset selection. We still need to be very deliberate that the datasets that we are selecting are (1) actually aligned with the tasks that we're trying to solve and (2) come from a sufficiently diverse and well-balanced distribution to make sure, you know, we're not unfairly, you know, we're not unfairly disadvantaging any sector of the innovation economy whether that, you know, disadvantage is along, you know, traditionally suspect classifications like demographics or even if, beyond demographics, right, even if it's the case that, you know, an AI model accidentally, you know, imposes different standards in the electrical arts versus mechanical. Even that distinction is something that we very much would want to avoid and that we can avoid by being very deliberate about the composition of datasets. [00:30:00] So, I think in short, you know, the accountability and the fairness issues associated with AI at large are present in the PTO context and that's something we think about all the time. Although we are insulated from a lot of the very, you know, worst issues as far as AI bias and, you know, AI safety goes, it's still incumbent upon us to really make sure that the datasets we select and the models, you know, we train on those datasets and the specific methodology of training those models is aligned with what we actually want to see from this technology. And we're, you know, we're committed to that every day. We think about that every day. And, you know, it really is a discussion that happens at the outset of any proposed AI initiative.

Fowler: [00:30:50] It's an interesting advantage that using specifically the patent dataset, as you know, being able to avoid hate speech because you're largely using professional writing for the most part. But I would say that there is probably a few things that are kind of weird in patent writing that doesn't exist in regular English. So, for example, a lot of patents will have boilerplate text in them saying like, this is what a computer is, or they'll have claims which are technically English, but also kind of not English because it's sort of a weird version of object-oriented English. That's the way I usually explain it to people. So, do you see these sort of [strangenesses 00:31:29] that are in patents and not really regular speech affecting the models in ways that you need to curate further?

Ma: [00:31:37] That's a fascinating point. And when you talk about, you know, the vagaries and perhaps the idiosyncrasies of patent drafting, I immediately think in my head what that means for the entropy of the predictions in a model. So there are certain domains of AI where, you know, predictions are just high entropy as a matter of course, as a matter of e-underlying data distribution, and there are certain domains where the predictions are comparatively lower in entropy and I would say that patent drafting is one of those domains with lower entropy just because there are a lot of, you know, stock phrases in terms of art if you will, on phrases of art, if you will, that, you know, tend to appear very formulaically in pretty predictable patterns ultimately, you know, make it less of a guessing game as to any particular word or phrase in a patent, right? Just because there's so much more structure in a patent than, say, in fiction writing that, you know, the entropy of a model that's predominantly trained on patent data will decrease just as a

function of, you know, natural language patterns and such. What that means for examination -- again this is not like a scientific attempt to elucidate exactly what happens on a mathematical level -- but just qualitatively, what we might expect as a result is that the model is very much able to handle things that we call in distribution, which means language that very much follows the tried and true patterns of patent drafting, but might do comparatively less well on things that are "out of distribution," so things that don't necessarily follow the patterns that, you know, you and I are used to seeing on a day-to-day basis in patent applications. And if we think about the implications, you know, let's think about which type of inventors perhaps might be most prone to sending applications that are out of distribution. I would venture without proof, you know, about fairly strong intuitions, that these classes of inventors perhaps tend to be more pro se types, independent inventors of sorts who might, you know, in drafting their own patent applications tend to use more sort of natural English and less sort of, you know, phrases of art constructs, if you will. And it's in those cases that we need to be especially sensitive to making sure that what our models do out of distribution do not, you know, work to the disadvantage of any discrete group of inventors, right, because we want to do that in general, of course. We want to make sure that no group of applicants is ever disadvantaged by an AI model, but, you know, when you think about the equities involved, you know, one area in which we want to be particularly cognizant of that need is in the independent inventor and pro se space because these people are already those, you know, are the sorts of people that need perhaps, you know, an extra guiding hand in navigating the PTO adjudication processes and, you know, we want to do as much as possible to make sure that, you know, the technology that we do use even if it's trained on predominantly, you know, patent applications that tend to follow the normal patterns, that those models are still able to, you know, effectively work on and work with these sorts of applications from independent inventors and other groups that might tend to be more sort of natural English out of distribution, if you will.

Glenn: [00:35:23] I'd like to ask you, since AI is the topic here, we have a program that we use when we're drafting patents. I won't name it because I don't want to make a commercial plug, but it has an art unit prediction model in it. It will read your claims and predict the art unit that the Patent Office will assign to it. I know that the Patent Office assigns art units when applications submit, at least they used to, by having a person in the mailroom read through the application and decide. Is the Patent Office going to be automating that process or have they automated that process?

Ma: [00:35:55] Great question, and one that for various reasons I won't get into the nitty gritty specifics of. But I think, you know, if you think about one of the tasks that AI has canonically just very well recognized as being suitable for, one of those tasks is the, you know, art of classification or categorization, right? So we see this even going back a decade to, you know, 2012 with a computer vision challenge called Image Net where, you know, there's I think 1.2 million images and you, the AI researcher, are tasked with sorting those images into one of a thousand different categories. That perhaps, you know, if you think about the

current wave of the deep learning accelerated AI revolution, that task was probably the first breakthrough result in the current wave of AI. You know, just this task of sorting images in a thousand different buckets. It's a bit more tricky of a task with English language in that English language tends to be a lot higher variance than, you know, the 256 by 256 pixel rosters that you might expect from an image. But ultimately, you know, as far as natural language goes, classifying natural language into discrete buckets is probably one of the easier tasks out there still. So as far as, you know, what's lower hanging fruit, what's less low hanging fruit, I think classification to those in the patent space and specifically in the patent AI space tends to be one that's regarded as perhaps on the lower hanging fruit side of things simply because, you know, we see all of these different examples of AI being used very, very effectively to sort things whether they be images or corpora of text or what have you into buckets. So we do not do that sorting at the art unit level, but we do sort patents, as you all might know, into classification codes. It used to be, you know, into something called the USPC, U.S. patent classification and now into the CPC schema or cooperative patent classification regime. And you know, you can think of AI being a very suitable tool to assign those CPC codes and then those CPC codes then being used in turn by internal USPTO processes to figure out who is the right examiner for any given application based on, you know, the needs of the office, based on the specific subject matter disclosed and, you know, other criteria that the office would seek to use to make an optimal match.

Glenn: [00:38:43] Going down the list a little bit, and I hope you don't mind. There's a question about does the PTO ever license that technology to others? And there's a subsidiary or corollary question to it is, does the PTO ever patent its own technology? I wonder if you could comment on that.

Ma: [00:38:59] Sure thing. Because our mandate is one of serving the public, we do not typically selectively license our technology to specific licensees. You know, ultimately, we are a fee funded agency. Any dollar that comes into the doors of the PTO is one that's paid by our stakeholders, so we want to make sure that any, you know, fruits of the labor derived from, you know, that revenue is directed toward the benefit of said stakeholders. So we won't, you know, selectively license what we're building, but we do release our technology and two - on a maybe more - but at least two channels. First, we often engage in partnerships with other government agencies or international IP offices where a portion of those partnerships will be an exchange of ideas and sometimes even of technology. And second, we often, when appropriate, seek to release what we're doing as far as technology development to the public at large so we can't do this with every technology product that we develop, you know, AI or otherwise, due to, you know, business needs and the exigencies of the office and, you know, the need for confidentiality as to certain matters. But where appropriate, we do seek to deliver useful technology to the patenting and trademarking public at large, and specifically to those who might be under-resourced or self-represented. And one recent example, actually, I guess two recent examples of this one and the, you know, just traditional technology space and one in the AI space. In the traditional

technology space, we recently launched a public version of our new and improved search infrastructure that, you know, in large part does resemble the internal search infrastructure that our examiners are using for their prior art retrieval and, you know, that we view as a very, very strong success story for developing technology in a manner that suits both the needs of our internal stakeholders as well as is adoptable for public use in some way, shape or form. So I think I, along with all my colleagues at the PTO, are very, very proud of that effort. On the AI side of things, we recently launched a beta version of what we call our inventor search assistant that uses AI retrieval technology not exactly in the same underlying system, but in the same spirit of the technology that we're building for examiners as far as AI goes and, you know, we're releasing that to the public. I forget the exact URL, but anyone who's interested can probably just search USPTO inventor search assistant on your favorite search engine and find it. But that tool is an example where we have specifically now thought outside the box of, okay, we're doing great things in the AI space. A lot of that development is directed toward internal needs, but how can we get a version of that in such a form that it is suitable for public use and it's specifically adaptable in a way that is intuitive to, you know, self-represented folks and other folks who are in most need of these tools rather than, you know, the typical sort of caricature of a patent search console with four windows tiled in such and such way and, you know, perhaps very dense, you know, tables of information, abbreviations all over the place. So with our inventor search assistant, we really thought from the ground up in terms of, okay, we want this to be useful to everyone, but most useful perhaps to self-represented folks and other folks who are in need of these sorts of resources. So given that, you know, initial goal, how do we design the product in a way that not only uses this technology, but does so in a intuitive interface and one that, you know, the everyday inventor can hope to decipher.

Fowler: [00:43:16] So I'm going to let you brag a little bit more about this search assistant thing. Can you go into a little bit more detail on what it does and what a member of the public who would encounter it would see when they operated it?

Ma: [00:43:29] Sure thing. So with the beta version of the inventor search assistant, which, you know, I want to give a shout out to one of my colleagues, Scott Beliveau 00:43:38, who, along with his team, now really laid the vision as well as executed the nitty gritty technical roadmap to make this a reality. So when I'm describing this work, I'm really just describing work that, you know, I'm really excited about but where credit is really due to, you know, one of my extraordinary colleagues in the office but with the inventor research assistant, the concept is, you know, you type in a natural, you know, language prompt in the search assistant that can now just be a short sentence, a short phrase or a longer excerpt from a draft specification that you might have. You know, you paste whatever that, you know, excerpt of text is into the box and click search and then it gives you a rank ordered list of relevant prior art. At least relevant publicly accessible prior art where the retrieval metric - the similarity score, if you will, is one not solely based on keyword match like a traditional search engine, but actually incorporates AI and natural language relevant metrics and making sure that now

the rank order list of prior art for whoever is using the inventor search assistant [00:45:00] is really directed to conceptually the thing that the excerpt that they pasted rather than just doing a, you know, frequency based standard information retrieval style keyword match. So the key, you know, the key insight or I guess the key ambition of the inventor search assistant is really to give folks a tool that can complement, you know, maybe their existing keyword search process that they might implement with another publicly available [inaudible 00:45:35] out there, either from the USPTO or from a third party, and allow them to dig one level deeper into things like conceptual similarity and, you know, just deviating beyond the traditional keyword paradigm.

Fowler: [00:45:49] Does that correlate at all to the tools that the examiners are actually using?

Ma: [00:45:53] Yes, it does. So examiners are still very much, you know, trained on keyword search and, you know, devising clever keyword searches augmented by all these fancy Boolean operators that I can't even remember off the top of my head. But along with that standard paradigm of examiners searching, we are building in augmentations to their search technology to give the same sorts of, you know, more conceptual similarity capabilities implemented via AI and natural language processing to help them augment their existing workflows. So I'm not going to make any promises like in three years or in five years we will have supplanted keyword search entirely. I think keyword searching is still very much going to be a bread and butter part of prior art retrieval. But I do think there is a lot of room in the spectrum between searching in keyword space and searching in concept space where we can make sort of intermediate, you know, step by step interventions to build and to bake in some of these more conceptual similarity capabilities. And, you know, we're always thinking, you know, in any sort of large organization beyond the core technology development, there's this question of change management, right? So you could build the best technology, you know, state of the art on the research frontier and, you know, seek to deploy it to examiners. But if you don't carefully reason about how you're going to get from step A to step B and what intermediate steps you're going to take along that journey, then, you know, by just replacing what everyone knows and loves and is used to with this thing that you certify based on information and belief is better, but that, you know, they haven't had a chance to really get to know and, you know, wrap their heads around, then that's not a recipe for success. So any time we're developing AI capabilities, it really has to be not only a exercise of raw technology development, building the latest and greatest in terms of AI models, algorithms, but also the incremental process whereby we start deploying these algorithms in a way that, you know, actually comports with examiner needs and is, you know, rolled out in a way that is adaptable by examiners and the public for those tools that we released to the public in a way that doesn't just completely disrupt how folks, you know, think about things.

Glenn: [00:48:23] You know, I've been patent practicing before the patent law offices for a long time, and I have a tremendous amount of respect for the people who work

in the patent office. I think they're consummate professionals. I think the patent office, by administering the patent system as well as they do, they spur innovation in this country. I also notice that the PTO has technology programs, IP programs and awards, and I list them, you know, and the questions: one was National Medal of Technological Innovation, USPTO National Inventors Hall of Fame, Collegiate Inventors Competition, National Science Technology Metals Foundation, National Academy of Inventors. So I think what I'm asking is, do you see the Patent Office as being very, very proactive not only in discharging its statutory obligation to examine patents, the patent system itself being a spur to innovation, but the patent office also itself being a spur to innovation, both in terms of these programs they have and the type of technology development they do.

Ma: [00:49:21] Michael, I'm really, really glad you brought up these programs because they highlight what I think is an often underappreciated role of the agency. So as you know, we, of course, have our constitutional mandate under Article 1, Section 8, under which we do the bread and butter work of adjudicating and issuing IP rights. And that work is very much about incentivizing innovation and creativity in our country, which I guess, as you've undoubtedly seen in your legal practices, unleashes our nation's inventive spirit toward solving crucial and sometimes existential world challenges. But here's the thing: we don't stop at incentivization at the USPTO. That's because people need, in the first instance, to know about the possibilities and the opportunities in the innovation economy before they're even able to participate in it, right? You can't play in a game that you don't even know about and so the agency's priority along this axis is, in short, to broaden the base of those who become scientists, engineers, inventors and innovative entrepreneurs. So it was a focus area for the agency, I think, going back at least to Director Michele Lee's tenure. Then, of course, Director Andrei Iancu who formalized some of the initiatives in this area under what's called the NCEAI or National Council for Expanding American Innovation. And of course, our current director, Ms. Cathy Vidal, has redoubled our commitment to this sort of broad based innovation under what we call the CI2 or Council for Inclusive Innovation. So I think when we're thinking about our efforts under these many outreach channels at the PTO, such as our five regional offices or our recognition programs for transformative innovators, or even our early career engagement efforts to foster the next generation of technologists. The name of the game here is illumination. So we're trying to shine a light on what we think is the limitless potential for folks to participate in the innovation economy, with the aim that our efforts will ultimately result in folks from all walks of life and all parts of our nation to hop on board. So I think, in short, most people think of what we do as incentivizing innovation, right? And that's an absolutely crucial part of our mission. I certainly won't deny that. But without illuminating the possibilities in the space and without fostering a broad base of new talent, these incentives won't necessarily reach all corners of this nation and spur innovation from, you know, everyone that we think has a role to play. So our outreach and recognition programs really aim to spread the good word that there is, you know, there's really never been a better time in the history of our nation to innovate and to create, whether it be in general or whether it be specifically in emerging fields

such as AI and machine learning and quantum computing or whatnot, you know, whatever your, you know, field of interest and expertise is, there's just really never been a better time than right now that in the year 2022, you know, to hop on board and to join the innovation economy. And what we do with all the initiatives that you describe is really just drive this message home to anyone who will listen.

Fowler: [00:53:03] Do you view yourself as a developer of technology in this sense, or is it a facilitator or how would you describe it?

Ma: [00:53:10] So I think if we go back to, you know, my categorization of the agency's role into sort of three hats, that of IP office that of AI and emerging technology contributor and that of, you know, consumer. I think our role, we do have a role as a developer of technology and it's probably when wearing the latter two hats, that of consumer and that of contributor. So I think at the USPTO, we are very deliberate about not building technology for the sake of building technology. So we don't want to just build cool, shiny things that, you know, just send the [garage 00:53:52]. A lot of what goes on in academia and industry is really unfettered exploration of the knowledge space and the idea space. And at the PTO, we love to see that and we love to support it as best we can. But going back to something I said earlier, we at the PTO are a fee funded agency. Where every dollar that comes in is one that ultimately our stakeholders pay to us. And as a result, we really have a responsibility to ensure that our own development efforts at the PTO are directed toward benefiting the innovation community. And so, you know, the specific areas in which we're working, such as search classification and some other areas, I don't think we've explored in depth like authentication of IP rights, customer assistance, especially to independent self-represented folks. Process automation and even AI and machine learning for security's sake. As you know, under 35 U.S. Code Section 122, we are actually required by statute to keep certain information submitted by our stakeholders confidential for certain periods of time. And, you know, in the current threat landscape of, you know, hackers, advanced persistent threats, whatnot, you know, security is at the forefront of the agency's agenda in AI. And machine learning has a really important role to play in terms of threat analytics, in terms of proactive system hardening and helping us achieve that statutory mandate of safeguarding inventors' information. And I think along all of these areas, if you look at our portfolio of investments, you will really see that they are targeted not just toward developing AI and learning technology in a sort of pie-in-the-sky manner, but really in a manner that really speaks to what we see as both the immediate needs and the sort of longer-term aspirations of the innovation community and the IP stakeholder community. So the short answer is yes, we most certainly are a developer of technology, but we [cabin 00:56:04] our role in that regard to that which will best help our stakeholders.

Fowler: [00:56:09] What do you view as the next target for development at the USPTO?

Ma: [00:56:16] That's a question I think about all the time that I might not have a super-satisfying answer to, simply because in the emerging, in the technology

space in general, in many federal agencies, technology products are delivered through a fairly standardized process of IT procurement, where the agency writes up a detailed statement of what we want. The agency sends that statement out and solicits bids. We get the bids. We select the most promising out of them, and hopefully all goes well. At the USPTO, one the challenges we face is that our domain is just so specialized. So there are two places where subject matter expertise and IP tends to go: either the private IP bar where you are at right now; or government agencies such as the USPTO. So when we put out a statement of work there might arise is sort of like lost in translation effect between us and the agency and the external audience with those statements out in the technology development community. So when we engage in tech development, especially in emerging tech, we really make sure to involve subject matter experts from the examining core and other divisions of the office to make sure that the technology muscle of external vendors is directed to the task of solving the most pressing challenges for patent and trademark administration. So, you know, part of it is that, you know, the aims of the agency will always be, you know, influenced in principle part by what those precise needs are, both from a stakeholder and from a staff perspective at any given point in time. But the thing - there's another facet that stems from this work being emerging technology specifically, and that's that emerging technology is just not mature, right? Emerging technology is perhaps the antonym of mature technology, so there's a ton of ambiguity and uncertainty any time you engage in a emerging technology project, and there's often this threshold question of mere existence that is, does the technology actually exist at all, given the current state of the art to really solve the problem we're trying to solve. So another wrench in this whole puzzle is that before developing an actual emerging technology product, we first need to find this what I call a proof of existence as to a potential solution. So, you know, this proof of existence problem really complicates any sort of long-term road mapping in emerging technology and AI and ML, because we now have aspirations that extend, you know, all the way to the moon as far as what we would like to see from emerging technology. But when push comes to shove, the only things that we will be able to execute on are things where we have solved that threshold question of finding the proof of existence because, you know, a lot of technology products, you know, you want to build a website with a login and a database. No, we don't need a proof of existence for that. We already know that that project can be done if we now invest sufficiently in building that sort of system with things like, you know, AI-based search with things like, you know, classification, AI for security and threat analytics. You know, we don't know if there are solutions and even conditioned on their existing solutions, we don't know precisely what they look like. So any-long term road mapping exercise, there needs to be a fair bit of room built in just for the inherent ambiguity as to any sort of emerging technology and that, [01:00:00] more than anything, really informs how we execute emerging technology at the USPTO. We don't, you know, we aren't really about these, you know, grand multi-year plans where by year-end we're going to have something that we can spell out in concrete terms. We are about fast iteration, about agile development, about moving forward quickly when, you know, when one avenue

doesn't seem fruitful to very quickly pivot to another avenue to solve the same problem that we're trying to solve and really as quickly as possible to find proofs of existence that suggest areas of promise that we can then invest further in.

Fowler: [01:00:39] I have a quick follow-up to this. So you would basically not characterize the PTO's innovation as sort of pie-in-the-sky, figuring things out. It's you want to identify a problem you think you can solve immediately and or at least and not necessarily immediately, but you think there's a feasible path there pretty quickly, right?

Ma: [01:01:04] Exactly. And because we, you know, we are charged with being good stewards of the resources that now are paid into USPTO coffers, we really are intentional about finding that proof of existence as expeditiously as possible, because if we're just, you know, milling about for many years on end, tackling the same problem, you know, without finding the proof of existence, then in some sense, we really have just transitioned to sort of more speculative, less directed, just sort of technology exploration, if you will. And that's something that's good in measured doses and in sort of a cabined, you know, in closely cabined covered portions of our technology roadmap. But it's not something that we want to predominate in our technology roadmap. So we really want to be deliberate about getting to the proof of existence quickly. And one strategy which we've been using more and more is actually to involve the public, to involve the AI research community across academia, research and even independent developers and data scientists to help us along this journey because the expertise for finding these proofs of existence oftentimes lie in that research community. So we already serve that research community by way of our work in the intellectual property space and now we're hoping that they, in turn, can help us in finding these technological proofs of existence and turn allow us to serve them better as we see a very symbiotic relationship between us and researchers as to now us helping them with the administration of the IP system and then them being able to help us by accelerating our technology development roadmap, especially in emerging technology fields.

Glenn: [01:02:56] Jerry, I want to thank you so much for taking time to meet with us today. It's been a very interesting discussion, and I very much appreciate your insights and sharing the role that the Patent Office is playing and not only examining patents, but in pushing the frontiers of technology forward as well.

Fowler: [01:03:15] This concludes this episode of the Perpetual Motion Podcast. Thank you for listening.

[END]