TOWARD A NON-DISPOSITIVE, HUMAN-FIRST AGENDA FOR PUBLIC SECTOR AI

BY

JERRY MA

Director of Emerging Technology and Designated Responsible Official for Trustworthy AI. U.S. Patent and Trademark Office. The views expressed in this article are the author's own. They do not necessarily represent the position of the Federal Government, the U.S. Department of Commerce, or the U.S. Patent and Trademark Office.

Visit **www.competitionpolicyinternational.com** for access to these articles and more!

Toward a Non-Dispositive, Human-First Agenda for Public Sector AI

By Jerry Ma

The current era of artificial intelligence ("AI") has engendered profound industrial transformation. Firms from social media to consumer finance are inextricably integrating AI into their core operations. Meanwhile, regulators and civil society grow increasingly wary of what they perceive as unaccountable algorithms deciding what media the public should see, what products they should be offered, and what contractual terms they deserve. And as governments begin to look toward AI to better serve citizens, such concerns translate readily — and often in intensified form — to the public sector. Governmental entities that focus on relentless automation, skilled workforce replacement, and metric optimization in their AI development agendas risk producing the same unaccountable outcomes as those already observed in the wild. But the public sector is not bound by the same imperatives driving private-sector AI development. Governmental entities have the option to adopt a non-dispositive, human-first AI agenda. This agenda is deliberate in scope but no less ambitious than those of private-sector AI pioneers. It recognizes the simultaneous limitations of standalone "black-box" AI and the incredible potential of AI technology to empower humans. It does not champion the deployment of closed-loop AI systems in dispositional contexts. But neither does it cabin AI's role to mere toy problems. Rather, this agenda calls for the measured integration of AI capabilities into human-driven domains — in short, creating AI that "rides shotgun" with human experts sitting in the driver's seat. The field of intellectual property administration is offered as an emerging case study in non-dispositive, human-first AI development.

01
INTRODUCTION

One could be forgiven, heading into 2022, for feeling deeply conflicted about the role of artificial intelligence ("AI") in society. Heralded by the advent of powerful deep learning algorithms and fueled by the proliferation of "Big Data", today's AI revolution has led to remarkable — sometimes bordering on unbelievable — advances in myriad fields. Recent AI breakthroughs, including in search and planning, structural biology, software development, and modeling the manifold modes of human expression,[2] are paradigmatic examples of a general principle: that advances in AI possess unmatched potential to improve productivity, unveil whole new domains of human endeavor, and help us better understand each other and the world we inhabit.

> ***One could be forgiven, heading into 2022, for feeling deeply conflicted about the role of artificial intelligence ("AI") in society***

Yet from this pageant of innovation arose unanticipated risks. A Twitter dialogue bot from a well-respected research lab started posting hate speech and calls for genocide mere hours after its launch.[3] Automated recommendations led new Facebook accounts straight to photos of abhorrent violence.[4] And beyond the media society consumes, AI algorithms have influenced the jobs promoted to different demographic groups,[5] produced credit scores that differ in accuracy between such groups,[6] and led to other dubious outcomes. By deploying "closed-loop" AI systems, which render determinations without the benefit of human intervention, private-sector AI pioneers have prioritized business efficiency over risk mitigation. Governments are just now catching up to industry with emerging approaches to AI regulation and oversight.[7]

Governments, though, are increasingly entering the AI business themselves. And the public sector is far from immune to the risks revealed by private-sector AI deployments.[8] Indeed, because governments largely rely on the same AI model architectures, training algorithms, software libraries, and computing hardware pioneered by industry, it might seem inevitable that governmental AI efforts are doomed to repeat the same types of mishaps as those already observed in the wild.

For governments to mitigate risks in private-sector AI only to produce the same risks through public-sector AI would be the ultimate study in irony. Fortunately, governments have acted to prevent this double standard by establishing ground rules for responsible public-sector use of AI. In the United States today, Executive Order 13960 ("Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government") sets out the broad requirements for the use of AI by federal agencies.[9] These requirements include attributes such as

2 While no enumeration of AI breakthroughs can hope to be comprehensive, refer to AlphaGo, MuZero, AlphaStar, OpenAI Five (search and planning); AlphaFold (structural biology); OpenAI Codex (software development); and Transformer, BERT, GPT-3, DALL-E (modeling modes of human expression).

3 Rob Price, "Microsoft is deleting its AI chatbot's incredibly racist tweets," *Insider,* https://www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3.

4 Sheera Frenkel & Davey Alba, "In India, Facebook Grapples With an Amplified Version of Its Problems," *New York Times*, https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html.

5 Kim Lyons, "Facebook's ad delivery system still has gender bias, new study finds," *The Verge*, https://www.theverge.com/2021/4/9/22375366/facebook-ad-gender-bias-delivery-algorithm-discrimination.

6 Edmund L. Andrews, "How Flawed Data Aggravates Inequality in Credit," *Stanford University Human-Centered Artificial Intelligence*, https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit.

7 Elisa Jillson, "Aiming for truth, fairness, and equity in your company's use of AI," *Federal Trade Commission Business Blog*, https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai; Food and Drug Administration, "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan," https://www.fda.gov/media/145022/download; European Commission, "Proposal for a Regulation laying down harmonised rules on artificial intelligence," https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

8 See, for example, Larson et al., "How We Analyzed the COMPAS Recidivism Algorithm," *ProPublica*, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

9 Exec. Order No. 13960, 85 Fed. Reg. 78939 (Dec. 3, 2020).

safety, accuracy, and transparency, among numerous other desiderata.[10]

But a question looms large: how can government go about pursuing these laudable goals in its day-to-day AI activities? Given industry's mixed experiences with AI, it seems probable that an unstructured, ad-hoc approach won't suffice. Governmental entities will need to adopt a consistent development agenda whose underlying principles affirmatively advance trustworthiness and accountability across the portfolio of AI activities.

This article offers one such agenda, which recasts AI's role in the public sector from that of decision maker to that of helpful assistant. It then illuminates this agenda within the context of a U.S. agency, proving that a focus on putting humans first rather than on automated disposition is wholly consistent with pursuing an ambitious, impactful, and responsible AI portfolio.

02

PRINCIPLES FOR A NON-DISPOSITIVE, HUMAN-FIRST AI DEVELOPMENT AGENDA

The agenda set forth in this article first concedes that AI techniques — especially the highly parameterized models that underpin deep learning — are alchemical experiments in data metamorphosis. They transmute a given input into a desired output, with mathematical vector spaces as the intermediate substrates of this mysterious process. Descriptively speaking, this transmutation could very well *appear* to implement some cognizable procedure. But inside the black box, this transmutation operates not in the space of procedural reasoning, but rather in the space of statistical dependency.

Thus, using closed-loop AI systems to administer public affairs is fraught with risk. How can a governmental entity ensure that an AI-generated prediction corresponds to an actual decision-making basis prescribed by law or regulation? This is an impossible task in all but the simplest problem settings. One cannot extract reasoned judgment from the thousand-dimensional vector spaces traversed by AI's formulaic operation. And with neither a sound justification for these types of systems nor an overriding private imperative to improve the bottom line, the public sector simply doesn't need to risk deploying closed-loop AI systems in dispositive settings.

Yet this observation does not foreclose governments from using AI — far from it. AI is a tool, much like word processing or email. That governmental entities wouldn't write an automated Outlook rule to dispose of public complaints doesn't imply that they should forego email entirely. And that AI is similarly ill-suited to dispositive use doesn't imply that it should be ignored within the public sector. The public sector must simply focus on the unique strengths of AI.

Turning to those unique strengths, AI is unmatched in its ability to detect higher-order relationships from data. Patterns that escape humans can be recovered *ex machina* with the right AI model architecture. Relationships that humans *could* discern at high cost can instead be analyzed — with no capital investment — on commodity cloud computing resources at mere cents and seconds per gigabyte. AI can connect the dots: thousands, millions, or billions of them. It just can't decide what to do with those connections.

What should governmental entities do when the human expertise they need is expensive and supply-constrained, while AI computing resources are cheap and plentiful? The answer isn't complicated: use AI to make human experts maximally effective. The following principles elaborate on this core precept:

1. Governmental entities should steer clear of deploying closed-loop AI systems to autonomously dispose of public matters.

2. Governmental entities should identify the informational and contextual needs of their expert workforce, toward determining whether and how AI systems can meet such needs more effectively than the status quo.

3. Governmental entities should survey the "blind spots" currently faced by experts and explore AI solutions that can support experts in uncovering those blind spots.

4. Governmental entities may consider the use of AI systems for clerical tasks that disproportionately consume experts' time, so long as such systems involve one or more steps in which the expert reviews how the clerical function was performed and intervene as needed.

Together, these principles call for AI to empower — rather than replace — human experts by exposing relevant information, suggesting unapparent avenues of investigation,

10 *Ibid*.

and freeing up focus from rote distractions. And adopting these principles in an AI development agenda ensures that human expertise, married with AI-driven insights and freedom from repetitive tedium, remains the linchpin of public administration.

03

AN EMERGING CASE STUDY: AI AT THE U.S. PATENT AND TRADEMARK OFFICE

Although a few domains — such as defense, national security, and social services — stand out within the popular conception of AI in government, opportunities to practice the foregoing principles abound throughout the public sector. In fact, any governmental entity whose operations rely on sound human judgment and subject-matter expertise can stand to benefit by developing AI through a non-dispositive, human-first approach. As an emerging case study of such an approach, we turn to the U.S. intellectual property system.

The U.S. Patent and Trademark Office ("USPTO"), an agency of the U.S. Department of Commerce, is charged with the administration of the United States patent and trademark regimes.[11] The USPTO's principal mission is to grant patents and register trademarks in furtherance of scientific progress and economic growth. The agency fulfills this mission by adjudicating patent applications, trademark applications, and related matters.

One might be surprised to learn of the USPTO's significance in framing the contours of AI within the U.S. Government. While intellectual property administration is but one of the government's myriad functions, it predominates in the ecosystem of federal administrative adjudication. Out of an estimated 12,800 Executive Branch adjudicators in the U.S. Government as of 2017, over 8,000 served within the USPTO as Patent Examiners, Trademark Examining Attorneys, Administrative Patent Judges, and Administrative Trademark Judges.[12] Thus, the agency's AI development ventures necessarily shape AI's role within a sizable share of the government's adjudicatory activities.

Why does the USPTO perform so much adjudication? Simply put, the cases are numerous and complex, and they're growing only more so as time marches on. In 1790, when the first Patent Act was enacted in three pages of statutory text,[13] a total of three U.S. patents were granted.[14] Their adjudication was a collateral duty of then-Secretary of State Thomas Jefferson.[15] Fast forward to today — when patent grants number over 300,000 and applications over 600,000 annually (with even more activity on the trademark registers),[16] when patent doctrine resides in an entire title of the U.S. Code along with an intricate tapestry of decisional law, and when inventions encompass everything from quantum computers to mRNA vaccines — and it becomes perhaps less astonishing that over 60 percent of Executive Branch adjudicators serve within the USPTO. Millions of person-hours per year are invested in the operation of our intellectual property system, and this investment will likely only increase with continued scientific progress and economic growth.

04

CLOSED-LOOP AI: THE ROAD NOT TAKEN

Against this backdrop, it's tempting to dream of closed-loop AI systems that can dispose of patent and trademark cases. A patent specification — the heart of a patent application that serves as an "instruction manual" of sorts for the invention — follows well-recognized styles and struc-

11 U.S. Patent and Trademark Office, Overview, https://www.uspto.gov/about-us/overview.

12 Administrative Conference of the United States, "Non-ALJ Adjudicators in Federal Agencies," https://www.acus.gov/sites/default/files/documents/Non-ALJ%20Draft%20Report_2.pdf.

13 1 Stat. 109–112.

14 U.S. Patent and Trademark Office, "U.S. Patent Activity Calendar Years 1790 to the Present," https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm.

15 "Overview," *supra* at 11.

16 "U.S. Patent Activity Calendar Years 1790 to the Present"; U.S. Patent and Trademark Office, "Summary of Performance and Financial Information," https://www.uspto.gov/sites/default/files/documents/USPTOFY21PARSUMMARY.pdf.

tures, with the scope of the patented matter (the "claims") described through particularly formulaic patterns. Trademarks lend themselves even more readily to use as AI inputs — often amounting to single words or short phrases. Such filings, at first blush, seem precisely like the type of content amenable to the dispositive application of modern AI techniques. Compile a dataset of past cases, train a predictive neural network, and Bob's your uncle — or so it would appear.

In reality, any experienced practitioner of patent or trademark law would immediately discount this approach as categorically unworkable. The reasons are innumerable, but consider just one example in the patent context. A now-canonical test for the patentability of a purported invention (dubbed the "*Alice*/*Mayo*" test) requires a patent examiner to:

1. Determine whether the invention concerns "an abstract idea, a law of nature, or a natural phenomenon."[17]

2. If so, determine whether the invention contributes enough beyond the mere abstract idea, law of nature, or natural phenomenon to constitute something "significantly more" — that is, an "inventive concept."[18]

One does not need to be learned in patent law to intuit that this test often becomes an incredibly nuanced judgment call. What precisely is an abstract idea? A law of nature? A natural phenomenon? Something "significantly more"? Given that the federal judiciary is still hard at work drawing these contours amid increasing scientific and technical complexity, the notion that any AI system could correctly perform this single test — let alone the countless other decisions that feed into a determination of patentability — is fantastical.

Yet purely *quantitative* evaluations of such an AI system would likely indicate that the system "works" to some extent, in that its results are at least better than random guessing. The raw accuracy figures for the system might even appear to suggest real promise in certain circumstances. For example, it wouldn't be astonishing to witness an AI system achieve something like "90 percent accuracy" on a dataset of *Alice*/*Mayo* determinations under some intelligible definition of accuracy. This is plausible because AI algorithms, especially those of the deep learning variety, are powerful detectors of high-order statistical dependencies. Certain types of inventions, certain flowchart diagrams, or even certain words in a patent specification could — from a purely descriptive standpoint — correlate quite well to rejections under *Alice*/*Mayo* or any number of other grounds. And if an AI system can be trained to pick up enough such correlations, there's nothing stopping it from reaching any given quantitative performance milestone.

Were patent adjudication a profit-motivated affair — with said profits tied solely to "accuracy," labor costs, and other top-line metrics — then an adjudicatory enterprise might very well decide to deploy closed-loop AI systems to dispose of cases. Maybe the enterprise would look to replace human adjudicators entirely. Or maybe the enterprise would retain a small adjudication corps to perform quality assurance. But in any case, the operation would plod along — maybe even at some facially impressive quantitative accuracy — with decisions being rendered at lightning speed and near-zero marginal cost.

Of course, there's no free lunch. The seeming efficiencies realized by a closed-loop AI approach would come at a great cost to those who rely on the faithful execution of the patent laws. Such an outcome would, in a nutshell, be wholly unaccountable to the stakeholders within the intellectual property ecosystem.

First, a procedure relying on closed-loop AI simply couldn't be credibly described as adjudication in any sense of the word. The AI system, rather than following any intelligible set of rules and standards, would simply attempt to separate the cases labeled "allow" from those labeled "reject" using whatever promising statistical relationships its training process could encode. A procedure can't claim to adjudicate cases according to the patent law of the United States if it isn't actually designed to implement any law whatsoever.

Second, because of closed-loop AI's inability to perform true adjudication grounded in the law, such a procedure would lack robustness to adversarial exploitation. Patent prosecutors — those who assist inventors in obtaining a patent — are held by regulation to an exacting standard of legal, scientific, and technical training.[19] Faced with a closed-loop AI system, these intelligent and innovative professionals would have little difficulty finding the combination of magic incantations that can reliably elicit a positive outcome. Applying for a patent would become a farcical endeavor in which applicants focus on discovering tricks to

17 U.S. Patent and Trademark Office, *Manual of Patent Examining Procedure* § 2106 (Oct. 2019).

18 *Ibid*.

19 37 C.F.R. § 11.7(a)(2)(ii); USPTO Office of Enrollment and Discipline, "General Requirements Bulletin for Admission to the Examination for Registration To Practice in Patent Cases Before the United States Patent and Trademark Office" (Dec. 2021).

play on the AI system rather than on discovering new and useful inventions.

Lastly, because closed-loop AI cannot produce an intelligible account of the determinative facts, law, and reasoning that drive any decision, such a procedure would be unconstructive in helping applicants reach a satisfactory outcome. At its ideal, patent adjudication is a collaborative process between examiner and applicant. Although the examiner may formally "reject" an application (or portions thereof) in a response to the applicant, the response is made in the spirit of educating the applicant on why the application is not in condition for allowance, as well as ways in which the applicant can correct the situation. Applicants, in turn, work with examiners in an intricate process of interview, reply, and amendment to address any pending issues. A dispositive AI system would not live up to this collaborative ideal — applicants, upon receiving a rejection, would be deprived of any meaningful guidance to advance in the patenting process.

> ***Yet purely quantitative evaluations of such an AI system would likely indicate that the system "works" to some extent, in that its results are at least better than random guessing***

The end product of patent adjudication is binary: allow or reject. Binary classification has been among the most amenable environments in which to deploy closed-loop AI systems. But using closed-loop AI for adjudication at the USPTO would be a fragile charade — one that the agency has rightly dismissed.

05

NON-DISPOSITIVE, HUMAN-FIRST AI AT THE USPTO

AI systems cannot perform the USPTO's core adjudicatory functions, yet AI still stands among the agency's foremost strategic priorities.[20] How can this be so?

The answer lies in the USPTO's adoption of a non-dispositive approach to AI development. The agency's ambitious AI program aspires to empower its technical and legal experts to make well-informed decisions, rather than to relieve them of decisional responsibility.[21] In this way, the USPTO can reap the benefits of today's remarkable AI capabilities without incurring the most severe risks to accountability posed by dispositive AI.

Within the patent sphere, the USPTO deploys AI in two principal contexts: search and classification. Because an invention is patentable only if it is sufficiently original,[22] examiners must adjudicate each application in the context of what has already been done before. But the space of what has already been done before is so vast that even several lifetimes of undirected research would fall short. Thus, for examiners to faithfully administer U.S. patent law, the USPTO must be able to provide them with means to quickly retrieve and analyze the most relevant prior work. Since today's AI technology can uncover subtle — even conceptual — relationships between millions of documents, AI is especially well suited to power the USPTO's next-generation search systems. AI-based search capabilities are already helping USPTO examiners better ascertain the landscape of prior work pertaining to each application.

Another area of great promise for AI is patent classification. USPTO examiners are scientific experts, but their expertise is concentrated in specific areas of art. For a patent application to be properly adjudicated, it must first be sent to an examiner whose expertise matches the subject matter of the invention. This presents another natural opportunity for deploying non-dispositive, human-first AI. Specifically, the USPTO is developing predictive AI that can make initial suggestions regarding the types of technologies to which an application pertains.[23] These initial suggestions can then be used to route applications to the examiners who can best adjudicate patentability. Of course, predictive AI will never be perfect, which is why examiners retain the ability to submit corrections and have applications redirected appropriately. And because every classification is ultimately seen by at least one examiner, humans remain firmly in con-

20 U.S. Patent and Trademark Office, "Artificial Intelligence," https://www.uspto.gov/initiatives/artificial-intelligence.

21 U.S. Patent and Trademark Office, "Artificial Intelligence tools at the USPTO," *Director's Forum: A Blog from USPTO's Leadership*, https://www.uspto.gov/blog/director/entry/artificial-intelligence-tools-at-the.

22 35 U.S.C. § 102 (novelty); 35 U.S.C. § 103 (non-obviousness).

23 For more information on the scientific and technical taxonomy used to classify patents, refer to the Cooperative Patent Classification scheme, https://www.cooperativepatentclassification.org/index.

trol of the overall classification process. In fact, by flagging erroneous classifications, human experts play a direct role in improving the underlying AI algorithms over time.

Similar search and classification requirements arise in the trademark sphere, with both the USPTO and the public in need of information about which trademarks already exist, which goods and services trademarks are used for,[24] and which visual design elements are present in trademark images.[25] Furthermore, applicants are required to include "specimens" — proof of use of the trademark in commerce — in certain trademark applications,[26] and the USPTO maintains constant vigilance toward attempted frauds upon the agency in the form of forged or altered specimen submissions. USPTO AI efforts are underway toward addressing all these challenges. Of course, the resulting tools won't be used for automated disposition of trademark matters. Rather, they will be offered to Examining Attorneys and other trademark professionals, who will operate these tools toward ensuring the accuracy and integrity of the U.S. trademark registers.

In short, the USPTO has deliberately constructed its AI development portfolio to put human experts first, with AI systems placed in important but circumscribed supporting roles. As the USPTO proceeds with its non-dispositive AI agenda, agency adjudicators will continue to faithfully administer the nation's intellectual property system — as they have for the past two centuries — armed with technical aptitude, legal expertise, *and* best-in-class AI tools.

06
CONCLUSION

Just a decade ago, governments largely viewed AI as exploratory research to be funded rather than as operational capabilities to be deployed. They certainly would have been hard-pressed to identify even a few feasible applications of AI technology in public administration. It was firmly industry's remit to demonstrate that the convergence of algorithmic innovation, hardware accelerators, and large datasets could result in unprecedented opportunities for real-world impact. And the results have been such that governments now pay rapt attention to AI's possibilities in public service.

But governments are also examining the many risks that have emerged from private-sector AI innovation, and civil society is in turn considering whether and how these risks can arise in the public sphere — where the stakes can be much higher. The recent adoption of broad "Trustworthy AI" guidelines for the public sector indicates that governments are aware of the need to mitigate these risks. Yet individual governmental entities must still bridge the gap between such guidelines and their practical AI development agendas. In doing so, they must navigate between two extremes — on one hand, forgoing the use of AI entirely, and on the other hand, trying to automate as many decisions and processes as can be identified — in the shadow of their specific legal, regulatory, and subject matter contexts.

As long as AI remains an alchemical affair, AI's remit must be carefully managed. Allowing a closed-loop AI system to dispose of public matters reduces such matters to rote mathematics. And without a robust bidirectional interface between the mathematics of AI and the space of procedural reasoning, AI fails to provide a credible substitute for the human judgment and expertise that currently undergird public administration.

Yet AI still has a pivotal role to play in the public sector. The same attributes that militate against dispositive AI systems render AI exceptionally suited to many supporting roles alongside humans. By harnessing the unique capabilities of AI to uncover intricate descriptive relationships across millions of data records, governmental entities can develop user-facing tools to retrieve relevant information, decipher large corpora of data, flag issues for further investigation, and yet more. As a result, both internal experts and public stakeholders can redirect their attention toward the tasks that benefit most from human expertise.

A non-dispositive, human-first AI agenda acknowledges that "artificial intelligence", despite its name, cannot itself provide the intelligence that good governance demands. But it also recognizes AI's comparative advantages — uncovering patterns, drawing connections, and doing so at machine speed and scale — and places those superpowers firmly in human hands. Under this agenda, AI provides the context and support for public servants to leverage their independent expertise and discretion toward sounder outcomes.

24 U.S. Patent and Trademark Office, *Trademark Manual of Examining Procedure* § 1402 (Jul. 2021).

25 U.S. Patent and Trademark Office, "Design Search Codes," https://www.uspto.gov/trademarks/search/design-search-codes.

26 U.S. Patent and Trademark Office, *Trademark Manual of Examining Procedure* § 904 (Jul. 2021).

Responsible, accountable, and impactful public-sector AI isn't a pipe dream. Instead of expecting machines to think and to decide for us, let's start building AI that better informs our own thinking.

> ***Just a decade ago, governments largely viewed AI as exploratory research to be funded rather than as operational capabilities to be deployed***

CPI SUBSCRIPTIONS

CPI reaches more than **35,000 readers** in over **150 countries** every day. Our online library houses over **23,000 papers**, articles and interviews.

Visit **competitionpolicyinternational.com** today to see our available plans and join CPI's global community of antitrust experts.

CPI COMPETITION POLICY™ INTERNATIONAL