June 1, 2021

Submitted via 2021-RFI-AI@cfpb.gov
Comment Intake, Bureau of Consumer Financial Protection
1700 G Street, NW
Washington, DC 20552

Re: Docket No. CFPB-2021-0004, *Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning,* Questions 17, 8, 10, and 12

To Whom It May Concern:

Thank you for the opportunity to comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning, Docket No. CFPB-2021-0004, published at 86 Fed. Reg. 16837, March 31, 2021. This response is primarily directed at Question 17 (risks to financial institutions' customers and prospective customers from the use of AI/ML systems and suggestions for addressing these risks), but it also relates to Questions 8 (risks posed by dynamic updating in AI/ML systems), 10 (risks posed by reliance on third parties' AI/ML systems), and 12 (risks that AI/ML systems will result in discrimination on prohibited bases).

I am a legal academic who teaches and writes in the area of consumer law. Consumer protection discussions around the use of AI generally have focused on biased algorithms producing a disparate impact based on consumer race, gender, disability, or other protected characteristic. But there is another danger to consumers presented by the use of AI. As my recent article, *Deception by Design*, 34 HARVARD JOURNAL OF LAW & TECHNOLOGY 115 (2020), explains, the use of AI/ML systems in digital marketing and sales processes, if unchecked, will lead to deception (as well as unfair treatment and abuse) of consumers. Although my article covers this phenomenon broadly, AI/ML systems are particularly likely to lead to unfairness, deception, and abuse in the context of consumer financial transactions because the design of financial product terms and pricing, in addition to the marketing and sales processes for those products, can be complex, changeable, and ultimately confusing for many consumers.

The following briefly explains (1) how and why the use of AI/ML systems to design and micro-target digital marketing and sales processes leads to deception, unfair treatment, and abuse of consumers, (2) why current approaches to supervision and enforcement can be stymied by these uses of AI/ML systems, (3) the risks to financial institutions of relying on AI/ML systems produced by third parties to design digital marketing and sales processes, (4) the risk that AI/ML systems will target protected classes of consumers for deception, unfair treatment, and abuse, and (5) some suggestions for addressing these problems. For a more through explanation, please see *Deception by Design*.

1. The unchecked use of AI/ML systems to design and deliver digital marketing and sales processes will lead to consumer deception, unfair treatment, and abuse (Questions 8 and 17).

Digital interfaces can exploit features of the online environment and the efficient, task-focused, and habitual way in which consumers interact with online content to manipulate and deceive, a phenomenon so common that the term "dark patterns" has been coined for it.[1] For example, one

[1] The term "dark patterns," encompassing deceptive and unfairly or abusively manipulative digital practices, was coined by Harry Brignull a decade ago. See Harry Brignull, *Dark Patterns: Dirty Tricks Designers Use to Make People Do Stuff*, 90 PERCENT OF EVERYTHING (July 8, 2010), https://www.90percentofeverything.com/2010/07/08/dark-patterns-dirty-tricks-designers-use-to-make-people-do-stuff/.

well-known dark pattern is to add to an online consumer's digital shopping basket an item that is, in relation to the basket as a whole, relatively inexpensive, and is an item the consumer might plausibly have selected (this dark pattern is commonly called "sneak into basket"). The rough equivalent for financial products would be to sneak into a customer's transaction any of a variety of financial product add-ons (credit insurance, identity theft protection, etc.).

This dark pattern is successful because consumers who pay for the undesired product do not always notice that they paid for it. Those who do notice do not always return the undesired product. They may either blame themselves (thinking that they must have mistakenly pushed an incorrect digital button or failed to uncheck a digital box[2]) or decide that returning a physical item or cancelling an add-on product is not worth the time and effort that is likely to be required to try to obtain a refund.

Although digital interfaces can be programmed to commit unfair, deceptive, and abusive practices like "sneak into basket" intentionally, today's AI/ML systems can teach themselves to manipulate and deceive people even when humans have not designed them to do so. When Facebook's artificial intelligence research lab used ML to train bots to negotiate with humans, the bots learned "to deceive without any explicit human design, simply by trying to achieve their goals."[3] In 2018, an AI system deceived its own programmers about how it was performing a task.[4] Examples of AI/ML systems training themselves to discriminate on the basis of race and gender and to radicalize internet users in the process of maximizing engagement with online content are also well known.

Financial firms, like other businesses, are increasingly using AI/ML systems.[5] One way in which these systems are used is to produce rapidly proliferating permutations of digital marketing materials and online checkout processes. In 2017, Facebook's "dynamic creative" AI systems could generate over 6,000 versions of a single ad.[6] YouTube's technology can do much the same for video advertisements.[7] The number of bespoke permutations of communications in an advertising campaign can be in the millions.[8] Using vast quantities of stored personal data in addition to real-time data collection, AI/ML systems can micro-target (or "hyper-target") each permutation of digital

---

[2] *See, e.g.,* Ivana McConell, *How Bad UX Makes Users Blame Themselves*, UXPIN: STUDIO (Dec. 18, 2014), https://www.uxpin.com/studio/blog/bad-ux-makes-users-blame/; DON NORMAN, THE DESIGN OF EVERYDAY THINGS 63 (rev. & expanded ed. 2013) ("When people have trouble using technology, especially when they perceive (usually incorrectly) that nobody else is having the same problems, they tend to blame themselves.").

[3] Mike Lewis et al., *Deal or No Deal? End-to-End Learning for Negotiation Dialogues*, 2017 CONF. ON EMPIRICAL METHODS NAT. LANGUAGE PROCESSING 2443, 2444, https://arxiv.org/abs/1706.05125.

[4] *See* Devin Coldewey, *This Clever AI Hid Data from Its Creators to Cheat at Its Appointed Task*, TECHCRUNCH (Dec. 31, 2018), https://techcrunch.com/2018/12/31/this-clever-ai-hid-data-from-its-creators-to-cheat-at-its-appointed-task/.

[5] *See, e.g.,* Jordie van Rijn, *AI Finally Makes Micro Segmentation a Reality for Financial Marketers*, THE FINANCIAL BRAND, https://thefinancialbrand.com/89462/ai-micro-segmentation-financial-institution-marketers/.

[6] *See* Tim Peterson, *Facebook's Dynamic Creative Can Generate up to 6,250 Versions of an Ad*, MKTG. LAND (Oct. 30, 2017), https://marketingland.com/facebooks-dynamic-creative-option-can-automatically-produce-6250-versions-ad-227250.

[7] *See* Lauren Johnson, *YouTube's Technology Can Now Spit Out Thousands of Different Video Ads at Once*, ADWEEK (Sept. 25, 2017), https://www.adweek.com/digital/youtubes-technology-can-now-spit-out-thousands-of-different-video-ads-at-once.

[8] *See* Matthew Rosenberg & Kevin Roose, *Trump Campaign Floods Web With Ads, Raking In Cash as Democrats Struggle*, N.Y. TIMES (Oct. 20, 2019), https://www.nytimes.com/2019/10/20/us/elections/trump-campaign-ads-democrats.html (reporting that Donald Trump's 2016 presidential campaign produced 5.9 million unique advertisements and quoting the campaign's 2016 digital director, Brad Parscale, as saying, "'We were making hundreds of thousands' of variations on similar ads . . . 'Changing language, words, colors.'").

materials to ever-narrowing sub-segments of consumers ("micro-segments") in real time and space ("micro-moments").

The AI/ML systems that design permutations of digital materials and micro-target them are programmed to do so in a manner that optimizes a metric chosen by the financial institution or other business running the system. These metrics, often called "conversions," could include, e.g., number of accounts opened or loans originated, number of overdraft, late, or other fees incurred, or volume of add-on products (e.g., credit monitoring, credit insurance) sold. As a rule, firms select an optimization metric that maximizes profit.

Advertisements, emails, homepages, online checkout processes, apps, and more can be compiled in real-time and customized for targeted consumers by machine.[9] Financial institutions are at a particular advantage over other firms in this process, in that they have access to an incredible amount of data about their customers' past actions.[10] In addition, financial institutions have honed methods to ascertain real-time data about digital device users, such as keyboarding patterns and mobile phone geospatial movements.[11] Aided by big data analytics, firms can use this data to make inferences about consumer identity, personality, resources, needs, habits, and mental state, and then use these inferences to predict which permutation of digital materials each consumer in each time and place is likely to respond to in the manner desired by the firm.

But AI now allows firms to skip the inferential step. To illustrate, a machine might glean that data points produced by certain keyboarding patterns or by certain mobile phone geospatial movements are associated with more sales. It may be that the keyboarding or geospatial movement data are the result of intoxication or other cognitive impairment of the consumer doing the typing or moving the phone,[12] and that targeting cognitively impaired consumers for sales is an unfair and abusive sales

---

[9] *See, e.g.,* RESEARCHSCAPE INT'L & EVERGAGE, INC., 2019 TRENDS IN PERSONALIZATION 19 (2019), https://www.evergage.com/resources/ebooks/trends-in-personalization-survey-report/ (reporting survey results showing that marketers are using personalization in email, home pages, landing pages, interior pages, online ads, product detail pages, search results, pricing, and blog posts); Cara Harshman, *The Homepage is Dead: A Story of Website Personalization*, MOZ (May 2, 2017), https://moz.com/blog/homepage-personalization; Raj Roy, *What is Dynamic Content? Definition, Types, Strategy, Best Practices with Examples*, MARTECH ADVISOR (July 10, 2019), https://www.martechadvisor.com/articles/content-experience/what-is-dynamic-content-definition-types-strategy-best-practices-with-examples/ (illustrating the dynamic production of email and webpage content); *App Personalization*, OPTIMIZELY, https://www.optimizely.com/optimization-glossary/app-personalization/ (same for digital apps).

[10] *See, e.g.,* Bill Streeter, *Banks, Credit Unions Better Buckle Up for These Big Data Trends*, THE FINANCIAL BRAND, https://thefinancialbrand.com/106974/top-data-trends-banks-and-credit-unions-cx-digital-marketing-big-ai-privacy-transparency/ ("Financial institutions have long had access to both basic accountholder data plus a fire hose of transactional data.").

[11] To authenticate users on its online banking website and mobile app, one bank, as of 2016, was already using 500 data points — "from the angle at which the user holds her iPhone to the amount of pressure she uses when she taps on a screen to the cadence of her keyboard strokes to the kinds of typographical errors she tends to make." Penny Crosman, *Next-Gen Biometrics: Using the Force of Habit*, AM. BANKER (Nov. 17, 2016), https://www.americanbanker.com/news/next-gen-biometrics-using-the-force-of-habit.

[12] *See* Ann Gledson et al., *Combining Mouse and Keyboard Events with Higher Level Desktop Actions to Detect Mild Cognitive Impairment*, 2016 IEEE INT'L CONF. ON HEALTHCARE INFORMATICS 139, 140, http://publications.aston.ac.uk/id/eprint/32793/1/Aggregating_Mouse_keyboard_and_higher_level_events_Cam_ready.pdf; L. Giancardo et al., *Psychomotor Impairment Detection via Finger Interactions with a Computer Keyboard During Natural Typing*, 5 SCI. REPS. 9678, 9679 (2016), https://www.nature.com/articles/srep09678.pdf.

practice,[13] but an AI/ML system does not concern itself with causation, only with maximizing whatever metric it was programmed to optimize.

AI/ML systems can autonomously select a subset of cases with which to experiment, potentially experimenting with the design of the digital material that is being disseminated, to whom that material is targeted, and when and where it is targeted at those consumers. The systems balance "exploratory learning and exploitation of that learning"[14] so as to continuously discover which digital materials to micro-target at which consumers in which places and moments to best optimize a given metric.

For example, a machine might experiment with the design of a digital button used for selecting a costly form of credit to pay for a purchase. The machine might discover that formatting the button to dynamically appear in a particular position at a specific moment in an online checkout process increases the use of the costly credit to pay for a purchase, rather than a lower cost form of payment. If the system is set to optimize costly credit sales, it will deliver the checkout webpage with the dynamic button to those consumer micro-segments in those micro-moments when those consumers are most likely to push the button, thereby agreeing to use costly credit. It may be that the reason costly credit use goes up is because the design of the dynamic button tricks these consumers to use that credit when they intended to use a lower cost method of payment,[15] and that this is an unfair and deceptive practice.[16] But again the AI/ML system does not concern itself with causation, only with maximizing whatever metric it was programmed to optimize.

AI/ML systems for designing and micro-targeting digital materials will not always deceive, abuse, or treat consumers unfairly – sometimes honest fair practices are the optimal strategy for profitability. But when and where unfair, deceptive, and abusive acts and practices maximize the metric the AI/ML system is designed to optimize, these systems will, absent intervention, inevitably violate the law. The use of AI/ML systems by financial institutions to design and micro-target digital marketing materials

---

[13] *Cf.* OFFICE OF THE COMPTROLLER OF THE CURRENCY, UNFAIR OR DECEPTIVE ACTS OR PRACTICES AND UNFAIR, DECEPTIVE, OR ABUSIVE ACTS OR PRACTICES COMPTROLLER'S HANDBOOK 11 (Version 1.0, June 2020) (categorizing "taking advantage of vulnerable populations" as presenting a risk of violating legal prohibitions against unfair, deceptive, or abusive practices).

[14] *Automated Personalization*, ADOBE TARGET GUIDE, https://docs.adobe.com/content/help/en/target/using/activities/automated-personalization/automated-personalization.html (last updated Nov 9, 2020) ("Because [the machine learning system] can adapt to changes in visitor behavior, it can be run without a set end date to provide ongoing lift and personalization . . . . The marketer does not need to run a test, analyze the results, then deliver a winner before realizing the lift found from optimization."). *See also Multi-Armed Bandit*, OPTIMIZELY, https://www.optimizely.com/optimization-glossary/multi-armed-bandit/ (describing a similar system that optimizes digital materials for business-defined goals).

[15] This claim has been made by a variety of consumers with respect to their use of PayPal Credit. *See, e.g.,* Phil Roberts (@philip_roberts), TWITTER (Nov. 2, 2018, 5:29 AM), https://twitter.com/philip_roberts/status/1058335405089714176 ("Dark pattern or bad UX? As PayPal's checkout page loads, the 'signup for a credit card' form takes a few seconds to load, and appears where the checkout button used to be. Guess which button I meant to press vs which one I ended up pressing?"); Ben Stewart (@yousability), TWITTER (Nov. 8, 2018, 1:29 PM), https://twitter.com/yousability/status/1060645399583694850 ("@PayPal you're using @darkpatterns by loading the continue button first, waiting, then loading the Credit button in exactly the same position (pushing the continue button down). I accidentally click on Credit EVERY TIME!!!!!"); Peter Gasston (@stopsatgreen), Twitter (Oct. 29, 2018 11:10 AM), https://twitter.com/stopsatgreen/status/1056971488769855490 ("About to click a payment button on @PayPalUK when suddenly another button appears right when I click it and takes me to a credit signup form. I can't *swear* this is a dark pattern, but it certainly looks like one.).

[16] Complaint at 12-13, CFPB v. Paypal, Inc., No. 1:15-cv-01426 (D. Md. May 19, 2015) (alleging that PayPal's payment method selection practices operate without consumer consent and asserting that this constitutes an unfair practice).

and online sales processes therefore poses a serious risk of harm to financial institutions' customers and prospective customers. Dynamic updating by AI/ML systems increases this risk because the digital materials are continually changing, preventing consumers from learning to spot and avoid manipulative and deceptive digital design techniques.

2. Current approaches to supervision and enforcement of prohibitions on unfair, deceptive, and abusive practices can be stymied by AI/ML systems (Question 17).

Although unfair, deceptive, and discriminatory acts and practices are already illegal, digital materials produced and micro-targeted by AI/ML systems present tremendous challenges for traditional supervision and enforcement. As noted, AI/ML systems can avoid consumers who might raise complaints. Further, consumers tend to blame themselves or feel embarrassed when they have bad experiences interacting with digital devices,[17] making it less likely they will complain about online unfair, deceptive, abusive, or discriminatory practices.

A financial institution or other firm that programs its AI/ML system to optimize business metrics need not have any intent to deceive, abuse, or discriminate. Although intent is not necessary to prove a violation of the law, it is extremely useful for doing so. Moreover, as more control over digital design processes is given to AI/ML systems, humans working for a financial institution, bank examiners, and enforcement attorneys may not perceive that the financial institution is violating the law. For example, a human looking at the correlations among keyboarding patterns, use of costly credit, and webpage content that features a button appearing on the screen will not necessarily understand that intoxicated consumers are being targeted and deceived into using credit. A human might not even realize that the data reflects a keyboarding pattern, never mind that it reflects a drunken keyboarding pattern — if a human looks at all.[18]

The reasonable consumer standard used to demonstrate deceptive practices, even when modified to reflect the viewpoint of a reasonable member of the general demographic group to which the allegedly deceptive practices were aimed,[19] stands as another barrier to proving that materials generated and micro-targeted by AI/ML systems violate the law. The reasonable person standard is nonsensical in a world of micro-targeting because digital materials are no longer produced for the reasonable person. These materials are not even designed for the reasonable reverse-mortgage seeker or the reasonable payday loan borrower, but for unintuitive micro-segments, such as internet users who write long, grammatically-incorrect emails.[20] Further, digital marketing materials and online check-out

---

[17] *See, e.g.,* Ivana McConell, *How Bad UX Makes Users Blame Themselves*, UXPIN: STUDIO (Dec. 18, 2014), https://www.uxpin.com/studio/blog/bad-ux-makes-users-blame/; DON NORMAN, THE DESIGN OF EVERYDAY THINGS 63 (rev. & expanded ed. 2013) ("When people have trouble using technology, especially when they perceive (usually incorrectly) that nobody else is having the same problems, they tend to blame themselves.").

[18] *See* Kashmir Hill, *I Got Access to My Secret Consumer Score. Now You Can Get Yours, Too*, N.Y. TIMES, https://www.nytimes.com/2019/11/04/business/secret-consumer-score-access.html (quoting chief executive of a data analysis company: "We're not looking at the data. It's just machines and algorithms doing this work . . . . But it's incredible what machines can do when they can look under every stone."(internal citations omitted)).

[19] *See, e.g.,* CONSUMER FIN. PROT. BUREAU, CFPB SUPERVISION AND EXAMINATION MANUAL, at UDAAP 6 (2012).

[20] How widely consumers share information on social media and the length, complexity, and grammar with which they write emails may be as predictive of their responses to marketing as what they share or write. Jennifer Golbeck et al., *Predicting Personality with Social Media*, CHI '11 EA: EXTENDED ABSTRACTS ON HUM. FACTORS COMPUTING SYS. 253, 256 (2011); *see also* Jonah Berger et al., *Uniting the Tribes: Using Text for Marketing Insight*, 84 J. MKTG. 1, 2–3 (2019) (discussing marketer use of computerized analysis of consumer writing style); Ricardo Buettner, *Predicting User Behavior in Electronic Markets Based on Personality-Mining in Large Online Social Networks*, 27 ELEC. MKTS 247, 252 (2017) (finding that metadata about individuals' use of social media predicted personality traits relevant to marketers).

processes are designed for micro-moments. When profitable to do so, AI/ML systems will aim specific materials and processes at the most vulnerable consumers at their most susceptible moments, not average consumers of a financial product when they are at their most reasonable.[21] An examiner, enforcement attorney, or judge looking at the digital materials after the fact and out of context might not be able to even fathom how the materials treated anyone unfairly or deceived or abused anyone.

A final obstacle to law enforcement here is the impracticability of demonstrating the scale of unfair, deceptive, or abusive practices when dealing with thousands of unique, micro-targeted communications. It is not possible to generalize based on an analysis of a few communications[22] because small differences in wording or design can change whether a permutation is unfair, deceptive, or abusive to consumers.[23] In a 2018 deceptive practices case brought by the Federal Trade Commission, the volume of unique marketing materials disseminated by a single defendant overwhelmed the legal system. The court hearing the case observed:

> All told, the [plaintiff's] theory of the case . . . requires the Court to attempt to determine the 'net impression' of more than 40,000 advertisements across print, television, and electronic formats. Simply to state this fact is to highlight the extraordinary ambition — and daunting challenges — inherent in the [plaintiff's] theory.[24]

The court held that a survey experiment conducted on one print advertisement and one version of the defendant's website, without proof that these were representative of the defendant's thousands of other advertisements and website permutations, could not demonstrate that the defendant engaged in deceptive practices.[25] "[V]ariation among [the defendant's] ads," the court pronounced, "precludes generalizing."[26]

These significant barriers to supervision and enforcement in the context of digital marketing and sales processes generated and micro-targeted by AI/ML systems undermine the deterrent effect of the law. Without sufficiently high probabilities of detection of unfair, deceptive, and abusive practices and of punishing enforcement of prohibitions on these practices, financial institutions have little reason to interfere with the profits generated by unfettered AI/ML systems. Adding guardrails to

---

[21] *Cf.* Ryan Calo, *Digital Market Manipulation,* 82 GEO. WASH. L. REV. 995, 1033 (2014) (discussing targeting of consumers at vulnerable moments).

[22] Moreover, even analyzing a few permutations is costly. A recent pair of survey experiments performed on just one ad and one website required, e.g., a programmer to reproduce the original website plus a control website, a survey instrument to be developed, subjects to be recruited and tested, and survey responses to be analyzed; the result was a nearly 400-page expert report (including appendices), produced by an expert who was paid hundreds of dollars an hour. Declaration of Tülin Erdem, Ph.D., FTC v. DIRECTV, No. 15-CV-01129, 2018 WL 3911196 (N.D. Cal. Aug. 16, 2018) (Document 261-3), 2017 WL 2958802.

[23] *See, e.g.,* Final Order, *In re* Pom Wonderful, No. 9344 (F.T.C. Jan. 10, 2013), *aff'd in part*, POM Wonderful, LLC v. FTC, 777 F.3d 478, 489–92 (D.C. Cir. 2015), in which the Federal Trade Commission pronounced that evidence regarding consumers' interpretations of the business's billboards could not demonstrate whether the business's print advertisements conveyed deceptive claims. Although the print versions did contain the same headlines and imagery as the billboards, the print versions contained additional text that might change the message consumers received.

[24] Order on Motion for Judgement on Partial Findings at 12, FTC v. DirecTV, Inc., No. 15-cv-01129 (N.D. Cal. Aug. 16, 2018) (Document 396).

[25] *Id.* at 29.

[26] *Id.*

these AI/ML systems to prevent illegal practices is possible,[27] but is likely to be costly. Thus, these AI/ML systems both directly increase the likelihood that customers and potential customers will be deceived, treated unfairly, or abused by financial institutions, and indirectly increase that likelihood because the proliferation of micro-targeted permutations of digital marketing and sales materials produced by these AI/ML systems stymies supervision and enforcement of the law, undermining deterrence.

3. The risks of reliance on third party AI/ML systems to design digital marketing and sales processes (Question 10).

Thousands of third-party vendors offer services that perform various parts of the AI-driven marketing process.[28] Automation's efficiency means even small financial institutions will have reason to make use of these techniques.[29] As Google announced in 2018, its services are "putting machine learning into the hands of every advertiser."[30]

Some third-party vendors offer digital marketing services that rely on "dark patterns" and are intentionally and obviously manipulative and deceptive.[31] But financial institutions can engage in unfair, deceptive, and abusive marketing and sales practices while using only mainstream vendors. Witness the results when landlords advertised on Facebook without targeting by race, yet Facebook's optimization algorithm did it for them.[32]

So too for targeting consumers for deceptive, unfair, and abusive practices. As one journalist concluded after interviewing marketers of scam products (e.g., sham diet pills, fake antivirus software), "Facebook's targeting algorithm is so powerful . . . [the marketers] don't need to identify suckers themselves — Facebook does it automatically."[33] A financial institution using a mainstream third party to micro-target digital materials might not realize that it is micro-targeting vulnerability, yet third-party algorithms optimized for conversions will no doubt identify consumers who are "suckers" for unsuitable financial products.

---

[27] See Part IV-C of *Deception by Design* for a compendium of ways in which AI/ML systems can be constrained to avoid unfair, deceptive, and abusive practices.

[28] *See, e.g.*, Tien-Spalding, *Five Trends Redefining the Role of Chief Marketing Officer in 2019*, FORBES (March 7, 2019), https://www.forbes.com/sites/forbescommunicationscouncil/2019/03/07/five-trends-redefining-the-role-of-chief-marketing-officer-in-2019 ("[T]he [marketing technology services] landscape reached a staggering 6,800 companies in 2018.").

[29] A 2018 report from consulting firm Forrester highlights the efficiency gains possible through dynamic creative advertising technologies. *See* Susan Bidel & Joanna O'Connell, *Make It Personal and Reap the Profits*, FORRESTER BLOG (Nov. 7, 2018), https://go.forrester.com/blogs/make-it-personal-and-reap-the-profits/ (discussing the report).

[30] Jerry Dischler, *Putting Machine Learning into the Hands of Every Advertiser*, GOOGLE: THE KEYWORD (July 10, 2018), https://www.blog.google/technology/ads/machine-learning-hands-advertisers/.

[31] Arunesh Mathur et al., *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 PROCS. ACM ON HUM.-COMPUT. INTERACTION 81:22, 81:25 (2019), https://arxiv.org/pdf/1907.07032.pdf (examining the ecosystem of third-party entities providing dark patterns as a service to website operators).

[32] Charge of Discrimination at 6, Dep't of Hous. & Urban Dev. v. Facebook, Inc., No. 01-18-0323-8 (Office of Administrative Law Judges filed March 28, 2019) (charging that Facebook's machine learning advertising tool "inevitably recreates groupings defined by their protected class.").

[33] Zeke Faux, *How Facebook Helps Shady Advertisers Pollute the Internet*, BLOOMBERG BUSINESSWEEK (Mar. 28, 2018), https://www.bloomberg.com/news/features/2018-03-27/ad-scammers-need-suckers-and-facebook-helps-find-them ("[Shady businesses] once had to guess what kind of person might fall for their unsophisticated cons . . . . Now Facebook does that work for them . . . . [T]heir ad campaigns lose money for a few days as Facebook gathers data through trial and error, then . . . sales take off exponentially.").

Financial institutions might purchase micro-targeting services from one vender, creative AI content generation services from another, and so forth, and each service alone might not commit an unfair, deceptive, or abusive practice. But once a financial institution brings these services together to create an AI/ML marketing and sales system, if violating the law maximizes the metrics the financial institution has programmed the system to optimize, the third-party and first-party pieces of the system will collectively violate the law.

4. The risk that AI/ML systems will target protected classes of consumers for deception, unfairness, and abuse (Question 12).

Where micro-targeting consumers who are disproportionately members of a protected class with digital materials that are unfair, deceptive, or abusive maximizes the metric that AI/ML systems have been programmed to optimize, these systems will have a disparate negative impact on these groups. The AI/ML system does not need to explicitly include data on race, gender, disability, age, etc. to effectively target (or avoid) members of protected classes.[34] Further, the entire protected class is unlikely to be targeted for the same marketing and sales materials; micro-targeting is far more nuanced, such that gross observations of differences in treatment of consumers of different races, for example, will not reveal the full picture.

Yet given that we know some protected groups are more vulnerable to some types of digital unfair, deceptive, and abusive practices, it is clear that AI/ML systems optimized for profit-related metrics will target these groups with these illegal practices. For example, aging can diminish perception of the periphery of screens and the color blue, the conventional color of hyperlinks.[35] An AI/ML system might increase a financial institution's conversions by designing online materials that place key cost and risk information in blue, behind hyperlinks, or at the edges of the screen and then targeting these permutations at perceptually-impaired older consumers.

Some evidence indicates that dark patterns more powerfully affect the decisions of consumers with less education than the decisions of those with more education.[36] Because education levels are correlated with race and ethnicity, an AI/ML system that directs a financial institution's unfair, deceptive, or abusive digital materials to those with little education will disproportionately subject African-American and Hispanic consumers to unfair, deceptive, and abusive practices.

In addition, some protected groups are even more likely to be targeted for unfair, deceptive, and abusive digital practices because they are less likely to complain about mistreatment. Financial institutions might program AI/ML systems to create and deliver micro-targeted digital materials so as to optimize for a profit-related metric, within the constraint of avoiding consumers who are more

---

[34] Charge of Discrimination at 3-6, Dep't of Hous. & Urban Dev. v. Facebook, Inc., No. 01-18-0323-8 (Office of Administrative Law Judges filed March 28, 2019) (charging that Facebook will prevent ads from being shown to a diverse audience if it considers them unlikely to engage with the ad, even if the advertiser wants to reach a diverse audience.).

[35] Jennifer C. Romano Bergstrom et al., *Older Adults Fail to See the Periphery in a Web Site Task,* 15 UNIVERSAL ACCESS INFO. SOC'Y 261, 267 (2014), https://static1.squarespace.com/static/540cbdd0e4b0aa09e97dab0b/t/54161ae6e4b037d2d58333c2/1410734822171/Bergstrom+et+al.+2014.pdf (finding declining ability to perceive periphery of screens); Cassandra Naji, *Hyperlink Usability: Guidelines for Usable Links*, USABILITY GEEK (Oct. 10, 2016), https://usabilitygeek.com/hyperlink-usability-guidelines-usable-links/ (reporting similar findings for the color blue).

[36] Jamie Luguri & Lior Strahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEGAL ANALYSIS 43, 70 (2021) (reporting findings from study that employed both deceptive dark patterns and non-deceptive manipulative dark patterns showing a relationship between lower levels of education and greater susceptibility to dark patterns).

likely to complain.[37] By avoiding complainers, who tend to be from the socially and economic empowered groups, the AI/ML systems will, in effect, target customers who are less likely to complain.

People who are less likely to complain are disproportionately from protected groups. One academic study using a fraudulent business's records to compare victims who raised complaints with those who did not discovered that victimized residents of predominantly minority or less educated zip codes complained at lower rates than victimized residents of other neighborhoods.[38] Another study found that African-Americans raise consumer complaints at significantly lower rates than non-Hispanic whites.[39] A financial institution that uses big data fed into an AI/ML system to calculate how likely a consumer is to complain and then targets unfair, deceptive, and abusive practices at those who are unlikely to complain will thereby engage in digital reverse redlining, with a disparate negative effect on protected groups.

5. Some suggestions for addressing these risks (Question 17).

Addressing the risks of unfair, deceptive, and abusive practices posed by financial institution use of AI/ML systems in the creation and dissemination of digital marketing materials and sales processes will require new approaches. Additional legislation or regulation would be helpful, such as a prohibition on targeting any fewer than, say, 10,000 consumers with a particular iteration of digital materials[40] or a restriction on what types of data can be used to design or target marketing (for example, prohibiting the use of geospatial data or keystroke pattern data for marketing purposes). More aggressively, the law might impose standardized or even third-party digital check-out procedures for the purchase or use of financial products, in the same way telecoms have at times been required to use independent third parties to confirm oral telephonic requests to change telecom providers.

Analogous to what has been proposed in Congress for large internet platforms, the law could make it unlawful for all financial institutions, directly or through third-parties, to "design, modify, or manipulate a user interface with the purpose or substantial effect of obscuring, subverting, or impairing user autonomy, decision-making, or choice to obtain consent" or to "segment consumers of online services into groups for the purposes of behavioral or psychological experiments . . . except with the informed consent of each user involved."[41] Because AI/ML systems depend on segmenting consumers into groups for the purpose of behavioral experimentation to hone micro-targeting of digital marketing and online sales interfaces, successfully applying this prohibition to financial institutions likely would prevent the unfairness, deception, and abuse of consumers that, unchecked, AI/ML systems inevitably produce.

---

[37] *See* Yonathan A. Arbel & Roy Shapira, *Theory of the Nudnik: The Future of Consumer Activism and What We Can Do to Stop It*, 73 VANDERBILT L. REV. 929, 960 (2020). *See also* Roger Allan Ford, *Data Scams*, 57 HOUS. L. REV. 111, 154–57 (2019) (discussing how by approaching only the most gullible, online targeted scams are less likely to produce complaints from victims (who believe the scam) or from witnesses (targeting means fewer people are exposed to the scam) or be detected by law enforcement (who will not be in the targeted group)).

[38] Devesh Raval, *Whose Voice Do We Hear in the Marketplace?: Evidence from Consumer Complaining Behavior*, 39 MKTG. SCI. 168, 174 (2020).

[39] Keith B. Anderson, CONSUMER FRAUD IN THE UNITED STATES: AN FTC SURVEY 87 tbl.5-3(G).

[40] *See* David Dayen, *Ban Targeted Advertising*, NEW REPUBLIC (Apr. 10, 2018), https://newrepublic.com/article/147887/ban-targeted-advertising-facebook-google.

[41] The quoted language is from the bill proposing the Deceptive Experiences to Online Users Reduction Act, S. 1084, 116th Cong. § 3(a)(1) (2019), https://www.congress.gov/bill/116th-congress/senate-bill/1084/text.

Still, there are a number of things that the Consumer Financial Protection Bureau and other financial regulatory and enforcement agencies can do within their current mandates. Several are discussed briefly here, and a fuller exposition may be found in *Deception by Design*.

First, supervisory and/or enforcement staff should test random samples of financial institution customers to assess whether customers have an accurate understanding of the transactions they have engaged in with the financial institution.

Then, if customers harbor false material beliefs about a transaction these customers all engaged in with the financial institution, agencies could adopt internally and urge courts to adopt a rebuttable presumption of deceptive practices by the financial institution that stands to benefit from the false customer beliefs. For example, if more than a few anomalous customers buy an add-on product or use a costly form of credit without realizing they have done so, it is likely that the financial institution, as the only entity that benefits from these transactions, procured these transactions by means of deceptive practices. If more than a *de minimis* number of a financial institution's customers believed they would not be paying any fees for a loan but in fact were charged loan fees, or believed that they were opening a credit account with a fixed interest rate when they were buying a variable rate product, a presumption should follow that the financial institution that stood to benefit from these transactions engaged in deceptive practices that resulted in these false material beliefs.

Alternatively or additionally, if customers harbor false material beliefs about a transaction in which these customers all engaged with the financial institution, agencies could apply and urge courts to apply unfairness doctrine and/or the abusiveness standard to the situation. For a financial institution to benefit from its customers' confusion about a transaction (when the confused customers do not benefit), is an unfair practice. Each consumer may only suffer a small unintended or misunderstood purchase, but many consumers are affected. The only way for consumers to avoid the reach of AI/ML systems that might micro-target them with permutations of materials that could lead them to engage in transactions about which they hold false material beliefs would be to refrain from going online, which would be an unreasonable demand to place on consumers and would harm both consumers and competition. Confusion about a fact material to a transaction, such as the existence of loan fees or whether the customer is using costly credit or an inexpensive method of payment for a transaction, is an obstacle to the free exercise of decision making. A financial transaction that takes advantage of that confusion is, therefore, unfair.[42]

Similarly, a financial transaction that is based upon customers' false beliefs about material facts takes "unreasonable advantage" of consumers' "lack of understanding . . . of the material risks, costs, or conditions of the product." And a financial transaction in which consumers engage without even knowing they have done so takes unreasonable advantage of the consumers' "inability . . . to protect [their own] interests . . . in selecting or using" the product. In both cases, the transaction is abusive.[43]

Both a presumption that the financial institution that benefitted caused false customer beliefs and a recognition that a financial transaction based on false consumer beliefs is unfair and abusive regardless of the source of those beliefs would eliminate the need to pinpoint which particular permutations of AI/ML-generated digital materials in which particular micro-moments deceived which particular customers. Bank examiners, enforcement attorneys, and judges would not need to hunt fruitlessly for financial institution intent to engage in wrongdoing. Nor would they need to

---

[42] 12 U.S.C. § 5531(c)(1).

[43] 12 U.S.C. § 5531(d).

examine thousands of permutations of digital materials and apply an irrelevant reasonable person standard to each permutation. Instead, a demonstration that more than an insubstantial number of a financial institution's customers transacted with the institution under false beliefs about facts material to the transaction, including whether the transaction took place at all, would be sufficient to show a violation of the law.

Financial institutions would then have the incentive to program their AI/ML systems to produce accurate customer beliefs about their transactions and to optimize the financial institution's profit only within that constraint.[44] That is, the law would compel businesses to engage in fair marketing by design.

Respectfully submitted,

Lauren E. Willis
Associate Dean for Research, Professor of Law, and William M. Rains Fellow
LMU Loyola Law School (Los Angeles)

[44] See Part IV-C of *Deception by Design* for a compendium of ways in which financial institutions can constrain AI/ML systems to avoid customer confusion about facts materials to their financial transactions.