The Ethics of Robo-Advising

R' Mordechai Torczyner – torczyner@torontotorah.com

What is Robo-Advising?

1. Hildebrand, Bergner, *Conversational robo advisors as surrogates of trust*, J of the Academy of Marketing Science Nov '20

Robo advisors have been praised as the next operating system in finance and the "new wealth management interface of the 21st century" (Andrus 2014). Robo advisors provide investment advice without the intervention of a human advisor. In short, robo advisors are digital interfaces that guide investors through an entirely automated process of investment advisory from assessing financial goals, evaluating consumers' risk profile, and ultimately managing the entire portfolio (Faloon and Scherer 2017; Gomber et al. 2017; Williams-Grut 2017). While discretionary input from consumers is possible, the key property is the fully automated process of risk assessment, asset allocation, and portfolio management, consistent with consumers' current financial situation, financial goals, and appetite for risk.

2. Abraham, Schmukler, Tessada, *Robo-Advisors: Investing through Machines*, The World Bank

To help with investment decisions, robo-advisors start by defining the investment strategy of each individual based on his/her investment goals and risk profile. Robo-advisors ask potential clients about the purpose of the investment and the time horizon. Robo-advisors offer investment strategies for a variety of goals, including retirement, saving for large expenditures, establishing a rainy day fund, or generating a stream of income to cover expenses. These questions are complemented with objective and subjective questions that evaluate a client's willingness and capacity to tolerate risk. Objective risk metrics can include a client's income and years to retirement. Subjective questions ask, for example, how the client would react to a market decline and how comfortable he/she is with fluctuations in the market (Lam 2016). To keep costs low and the process simple, clients' assessments are conducted using standard online short questionnaires. Based on these two dimensions, robo-advisors use automated algorithms to make recommendations on how to allocate funds across different types of assets…

Although straightforward and time-saving, robo-advisors might not be able to know clients as well as human advisors do through multiple interactions, tailored questions, and closer relationships. "One-size-fit-all" questionnaires might be too simple and narrow to provide a complete overview of a client's financial situation and his/her needs. Furthermore, these questionnaires assume that individuals with a similar risk profile would provide the same answers to the same subjective questions, which might not necessarily be true (Deutsche Bank 2017). Robo-advisors also lack other important aspects of a client-advisor relation, such as helping clients define their financial goals, counseling during market downturns, or dealing with possible changes in their lives (Accenture 2015). Furthermore, limited risk-assessment might not provide a complete overview of a client's overall financial condition. Robo-advisors might not ask about a client's other investments (such as pension funds and real estate), future expenses, potential liabilities, spouse's financial condition, or insurances purchased, among other information (FINRA 2016). If robo-advisors act on partial information, they might not provide optimal recommendations…

3. Lisa Beilfuss, *The Future Robo Adviser: Smart and Ethical?* WSJ 6/19/18

Today's robo advisers aren't really all that intelligent, says Tucker Balch, professor of interactive computing at Georgia Institute of Technology. They're essentially simple programs doing what human advisers do, like making trades and rebalancing clients' portfolios—just much faster and more frequently. Eventually, that will become the baseline expectation of every investor, he says, which will push firms to use AI to try to differentiate themselves from competitors. Portfolios managed through robots likely will stretch beyond index-tracking funds into more active approaches that aim to spot opportunities in, say, specific stocks.

Some robos are expanding their offerings already. Wealthfront in March added a higher-cost fund that uses derivatives to replicate a popular hedge-fund strategy known as risk-parity. Wealthfront and others also offer smart-beta funds, which weight stocks by factors other than traditional market capitalization.

4. A guide to the best robo-advisors in Canada for 2021

https://www.moneysense.ca/save/investing/best-robo-advisors-in-canada/

5. Baker, Dellaert, *Regulating Robo Advice Across the Financial Services Industry*, Faculty Scholarship at Penn Law (2018)

At a minimum in our opinion, honesty means making only true statements about the products, the advisor's compensation, and anything else that is relevant to the products, the advice, and the purchase process, and honesty should also include accurately describing the basis for any recommendations, making any common sense disclosures that might be needed to correct a misimpression that the advisor is considering all of the products in the market if the advisor is not doing so, disclosing the existence of any compensation or other arrangements that might have the potential to bias the advice in a way that is not consistent with consumer's interests, and providing advice that is not biased in that manner.

6. Steven McCarty, *Human advisors vs. robo-advisors: Will ethics trump compliance?* ThinkAdvisor 4/17/15

Since computers standardize procedures, content and interactions, they effectively banish promissory lanaguage, unapproved sales techniques and materials, and misrepresentation (assuming, of course, that robo-advisor investment experts and software engineers aren't Madoffs in disguise!).

7. Mercadante, *4 Best Robo-Advisors for Socially Responsible Investing*, Money Under 30 11/26/20

It can be extremely difficult for an individual investor to develop a portfolio comprised of SRI compliant companies. There are very few companies that are 100 percent compliant. What SRI attempts to do is to invest in companies that are more compliant than industry average. As well, with companies being so diversified with both product lines and geography, it can often be difficult to identify the most appropriate companies.

There have been mutual funds based on SRI investing for decades. But robo-advisors offer investors an opportunity to invest in a portfolio of either funds or individual stocks that generally meet SRI requirements. In fact, it's possible to have a complete balanced portfolio with a robo-advisor that offers an SRI investment option.

<u>Our Questions</u>

1) In early 2021, a Reddit forum of retail investors upended the stock market via herd investing in Gamestop and other heavily shorted companies. In early 2023, a robo-advisor equipped with advanced machine learning algorithms and controlling billions of dollars in investments replicated elements of the 2021 strategy on a far greater scale, driving numerous hedge funds out of business in an illegal "pump and dump" scheme. Is there any liability for human beings in this case?
2) In 2022, Devious Investments LLP develops Cindy, a new, "conversational robo-advisor" which can mimic human conversation. They pretend Cindy is human, believing that investors will respond better this way, but they keep the fees at the low level normally charged for robo-advisors. Is this ethical?
3) Make Money Now LLP controls a robo-advisor which manages thousands of portfolios. The firm is approached by Cambran University, which wants to use the dataset from these portfolios for their artificial intelligence development program. May Make Money Now LLP sell the dataset?
4) Self Dealers LLP designs a robo-advisor with a preference for their own ETF's. Must they disclose that preference to the public?

<u>Case 1: The Renegade Robo Advisor</u>

8. J.K.C. Kingston, *Artificial Intelligence and Legal Liability*

Perpetrator-via-another. If an offence is committed by a mentally deficient person, a child or an animal, then the perpetrator is held to be an innocent agent because they lack the mental capacity to form a *mens rea* (this is true even for strict liability offences). However, if the innocent agent was instructed by another person (for example, if the owner of a dog instructed his dog to attack somebody), then the instructor is held criminally liable (see [4] for US case law).

According to this model, AI programs could be held to be an innocent agent, with either the software programmer or the user being held to be the perpetrator-via-another.

Natural-probable-consequence. In this model, part of the AI program which was intended for good purposes is activated inappropriately and performs a criminal action. Hallevy gives an example (quoted from [5]) in which a Japanese employee of a motorcycle factory was killed by an artificially intelligent robot working near him. The robot erroneously

identified the employee as a threat to its mission, and calculated that the most efficient way to eliminate this threat was by pushing him into an adjacent operating machine. Using its very powerful hydraulic arm, the robot smashed the surprised worker into the machine, killing him instantly, and then resumed its duties…
Direct liability. This model attributes both *actus reus* and *mens rea* to an AI system.
It is relatively simple to attribute an *actus reus* to an AI system. If a system takes an action that results in a criminal act, or fails to take an action when there is a duty to act, then the *actus reus* of an offence has occurred.
Assigning a *mens rea* is much harder, and so it is here that the three levels of *mens rea* become important. For strict liability offences, where no intent to commit an offence is required, it may indeed be possible to hold AI programs criminally liable. Considering the example of self-driving cars, speeding is a strict liability offence; so according to Hallevy, if a self-driving car was found to be breaking the speed limit for the road it is on, the law may well assign criminal liability to the AI program that was driving the car at that time.

9. Rabbi Yosef Karo (16th century Turkey/Israel), Shulchan Aruch Choshen Mishpat 418:7
השולח את הבעירה ביד חרש, שוטה וקטן, פטור מדיני אדם וחייב בדיני שמים. במה דברים אמורים? כשמסר להם גחלת וליבוה... שדרך הגחלת להכבות מאליה קודם שתעבור ותדליק. אבל אם מסר להם שלהבת, חייב, שהרי מעשיו גרמו.
One who sends a fire in the hand of a *cheresh*, *shoteh* or minor, is exempt by human law, and liable by heavenly law. When is this true? When one gave them a coal and they increased it… for a coal naturally is extinguished on its own before it can travel and ignite elsewhere. But if he gave them a flame, he is liable, for his deeds caused it.

10. Rabbeinu Asher (Rosh, 13th-14th century Germany/Spain), Responsum 101:5
שאלה ילמדנו רבינו חתן שיוצא מחופתו ודרך הוא שיוצאין עמו רעיו וקרוביו חוץ לעיר, וכך עשו... ורץ אחד בסוסו ופגע בפרד שהחתן רוכב עליה הכאה גדולה ובא לו הפסד לחתן... וטוען המזיק שמאחר שברשות הרבים היה שכל העולם רשאין ללכת בו אינו חייב לשלם שהיה לו לשמור עצמו...
תשובה: מה שטוען המזיק שהוא פטור לפי שהיה ברשות הרבים והיה לו רשות לרוץ והיה לו לניזק לשמור עצמו לאו טענה היא שאין לו לרוץ ברשות הרבים אפילו אדם ברגליו אלא כדי שיכול לעמוד כשירצה. כדאיתא בב"ק (לב א)...
ומזיק בגופו הוא כיון שהוא רוכב על בהמה והזיק בגוף הבהמה או באוכף שעליה כמזיק בגופו דמי...
Question: Teach us, our master: A groom left his *chuppah*, and normally his friends and relatives accompany him out of the town, and so they did… And one sped with his horse and struck the mule that the groom was riding, a great blow, causing a loss for the groom… The vandal claims that since everyone may travel in a public area, he need not pay, for the groom should have protected himself…
Answer: The vandal's claim that he is exempt because it was in a public area and anyone may speed there, and the victim should have protected himself, is not a valid claim, for one may only speed in a public area – even on foot – such that one could halt when he wished. This is in Bava Kama 32a…
And he is considered as one who has harmed with his own person [as opposed to his property], since he rides the animal and causes harm with the body of the animal or the saddle that is upon it. He is like one who harms with his own body…

11. Rabbi Yaakov Yeshayah Blau (20th-21st century Israel), Pitchei Choshen III 3 (5)
נראה שאף כשהפסיק ללחוץ על דוושת הדלק והמנוע ממשיך להפעיל ולהוליך את המכונית שחשוב ככוחו, וכל פגיעה של מכונית הנהוגה בידי אדם ה"ז ככוחו...
It appears that even when he stops pressing the gas pedal, and the engine continues to function and drive the car, it is considered his own force, and any time a car driven by a person strikes someone, it is [the driver's] force…

12. Rabbi Yosef Zvi Rimon (21st century Israel), *The Next Frontier in Jewish Law: Artificial Intelligence*, Jewish Action
Let's assume, for example, that a self-driving car causes harm to property—or, G-d forbid, kills someone. Although there may have been a human in the car, an autonomous system—AI—was in full control of the vehicle. Who is responsible? The "driver"? The programmer? The owner of the vehicle? The company testing the car's capabilities? The manufacturer?
RR: I was actually asked this very question after a self-driving vehicle did indeed kill someone. According to a *teshuvah* of the Rosh…

With regard to a self-driving car, however, the car is in control, and therefore the halachic category of *adam hamazik* does not apply; rather, the category of "*mamon hamazik*," property that causes damage, applies. This is similar to a case where a bull gores a person and injures him, which is classified as "*mamon shehizik*." Assuming the autonomous car has the status of *mamon shehizik*, who is going to take responsibility? It would seem that the owner of the car would have to take responsibility, since the car is his property. However, unlike the cases of *mamon hamazik* discussed in the Gemara, the owner of the vehicle could argue, "Why am I to blame? There is someone out there who programmed the car." One could counter that as the owner, he was obligated to ensure that the programmer's work was up to standard, and if he did not do so and his property caused harm, he is accountable.

The underlying question here is how halachah defines the basis of the owner's responsibility for damage caused by his property. Is an owner inherently responsible for damage caused by his property, unless there are circumstances beyond his control? If this is so, he would be responsible for damage caused by a self-driving car, unless he can prove there were extenuating circumstances. Or is he only responsible for damages that were caused due to negligence on his part, e.g., he didn't take sufficient security measures to ensure his animal or property would not cause damage? In that case, it would have to be determined that there was some degree of negligence.

13. Rabbi Yosef Karo (16th century Turkey/Israel), Shulchan Aruch Choshen Mishpat 420:25

צעק באזנו וחרשו, פטור מדיני אדם וחייב בדיני שמים.

One who screams in the ear of another, deafening him, is not liable by human law, but is liable in heavenly law.

Case 2: The Clandestine Robo Advisor

14. Hildebrand, Bergner, *Conversational robo advisors as surrogates of trust*, J of the Academy of Marketing Science Nov '20

The critical feature of conversational robo advisors is their capacity to take turns during the initial onboarding phase. Such turn-taking mimics natural iterations in human-to-human conversations which have been shown to act as an inherent trust-building mechanism (Bickmore and Cassell 2001)…

The process of turn-taking is often further governed by both verbal and non-verbal social cues, indicating active listening such as providing trivial acknowledgements of what the conversation partner just said or implicit signals to indicate whether the speaker is ready to yield the turn or whether an answer is expected from the listener (Wiemann and Knapp 1975). This back-and-forth communication protocol is an essential trust-building mechanism in human-to-human interactions. Indeed, even trivial acknowledgments or interludes of "small talk" can signal greater involvement and understanding from the side of the interaction partner and build stronger rapport (Bickmore and Cassell 2000; Cappela 1985)…

Thus, we expect that the inherent turn-taking capacity of conversational as opposed to non-conversational robo advisors enhances affective levels of trust toward the robo advisor. Affective trust is a distinct measure of relational trust between two parties, differing from other forms of trust (such as cognitive trust, which focuses on the objective assessment of competence and quality dimensions of an interaction partner; Johnson and Grayson 2005). Affective trust is a more emotional, subjective dimension of trust, linked to the social nature of a relationship. Taken together, we expect that the turn-taking capacity of conversational robo advisors increases perceptions of affective trust relative to non-conversational robo advisors.

15. Talmud, Gittin 14a

הנהו גינאי דעביד חושבנא בהדי הדדי פש חמש איסתרי זוזי גבי חד מנייהו אמרי ליה יהבינהו ניהליה למרי ארעא באפי מרי ארעא וקנה מיניה לסוף אזל עבד חושבנא בין דיליה לנפשיה לא פש גביה ולא מידי אתא לקמיה דרב נחמן א"ל מאי איעביד לך... הא קנו מינך א"ל רבא ואטו האי מי קאמר לא יהבינא דליכא גבאי קאמר א"ל א"כ קנין בטעות הוא וכל קנין בטעות חוזר

A group of vegetable merchants calculated together, and five coins remained with one of them. The others told him, in front of the farmer, to give it to the farmer; they performed a *kinyan*. In the end, the merchant made a personal calculation and nothing remained with him. He came to Rav Nachman, who asked, "What can I do for you?... The farmer performed a *kinyan*!" Rava said, "He didn't say 'I won't give'; he said 'I don't have any!'" Rav Nachman replied, "Then it was a mistaken *kinyan*, and mistaken *kinyanim* are reversed."

16. Talmud, Bava Metzia 80a

המוכר פרה לחבירו ואמר לו "פרה זו נגחנית היא, נשכנית היא, בעטנית היא, רבצנית היא," והיה בה מום אחד וסנפו בין המומין, הרי זה מקח טעות. מום זה ומום אחר אין זה מקח טעות.

If one sells a cow and says, "This cow gores, bites, kicks and sprawls," and it only has one of those defects, which he included among these defects, that is grounds for claiming it was an erroneous purchase. If he names this defect and one other, it is not an erroneous purchase.

17. Mishnah Nedarim 3:1 (20b-21a)

נדרי זירוזין כיצד? היה מוכר חפץ ואמר 'קונם שאיני פוחת לך מן הסלע' והלה אומר 'קונם שאיני מוסיף לך על השקל', שניהן רוצין בשלשה דינרין

What are "vows of goading"? He was selling an item and he said, "May this be consecrated if I would sell for less than a *sela*," and the other said, "May this be consecrated if I would offer more than a *shekel*," both of them intend to meet at 3 *dinar*.

18. Tosafot (12th-13th century Western Europe), Nedarim 21a

מתחלה כשנודרים היו רוצים בג' אלו שהיו ואנן סהדי דדעתן לכך הלכך לאו נדר הוא כלל

From the start, when they vowed they wanted 3; we can testify that they intended this. Therefore, it is no vow at all.

Case 3: Use of the Data

19. Anthony Tockar, *Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset*, Neustar Research 9/15/14

There has been a lot of online comment recently about a dataset released by the New York City Taxi and Limousine Commission. It contains details about every taxi ride (yellow cabs) in New York in 2013, including the pickup and drop off times, locations, fare and tip amounts, as well as anonymized (hashed) versions of the taxi's license and medallion numbers. It was obtained via a FOIL (Freedom of Information Law) request earlier this year and has been making waves in the hacker community ever since.

The release of this data in this unalloyed format raises several privacy concerns. The most well-documented of these deals with the hash function used to "anonymize" the license and medallion numbers. A bit of lateral thinking from one civic hacker and the data was completely de-anonymized. This data can now be used to calculate, for example, any driver's annual income. More disquieting, though, in my opinion, is the privacy risk to passengers. With only a small amount of auxiliary knowledge, using this dataset an attacker could identify where an individual went, how much they paid, weekly habits, etc. I will demonstrate how easy this is to do in the following section….

I spent some of the most riveting hours of my professional career searching through images of "celebrities in taxis in Manhattan in 2013" to find enough information to identify the correct record in the database. I had some success – combining the below photos of BC and JA with some information from celebrity gossip blogs allowed me to find their trips, which are shown in the accompanying maps.

In BC's case, we now know that his cab took him to Greenwich Village, possibly to have dinner at Melibea, and that he paid $10.50, with no recorded tip. Ironically, he got in the cab to escape the photographers! We also know that JA got into her taxi outside her hotel, the Trump SoHo, and somewhat surprisingly also did not add a tip to her $9 fare. Now while this information is relatively benign, particularly a year down the line, I have revealed information that was not previously in the public domain. Considering the speculative drivel that usually accompanies these photos (trust me, I know!), a celebrity journalist would be thrilled to learn this additional information…

But OK, perhaps you're not convinced. After all, this dataset is (thankfully) not real-time. How about we leave the poor celebrities alone and consider something a little more provocative. Larry Flynt's Hustler Club is in a fairly isolated location in Hell's Kitchen, and no doubt experiences significant cab traffic in the early hours of the morning. I ran a query to pull out all pickups that occurred outside the club after midnight and before 6am, and mapped the drop-off coordinates to see if I could pinpoint individuals who frequented the establishment. The map below shows my results – the yellow points correspond to drop-offs that are closely clustered, implying a frequent customer…

Examining one of the clusters in the map above revealed that only one of the 5 likely drop-off addresses was inhabited; a search for that address revealed its resident's name. In addition, by examining other drop-offs at this address, I found

that this gentleman also frequented such establishments as… Using websites like Spokeo and Facebook, I was also able to find out his property value, ethnicity, relationship status, court records and even a profile picture!

20. Talmud, Yoma 4b

מניין לאומר דבר לחבירו שהוא בבל יאמר עד שיאמר לו לך אמור שנאמר "וידבר ד' אליו מאהל מועד לאמר"

How do we know that one may not repeat that which is he told, until he is told, 'Go tell'? "And Gd told Moshe in the Tent of Meeting, to tell [the Jews]."

21. Rabbi Moses Maimonides (12th century Egypt), Mishneh Torah, Hilchot Deiot 7:5-6

המספר דברים שגורמים אם נשמעו איש מפי איש להזיק חבירו בגופו או בממונו ואפילו להצר לו או להפחידו הרי זה לשון הרע... כל אלו הם בעלי לשון הרע שאסור לדור בשכונתם וכל שכן לישב עמהם ולשמוע דבריהם

Statements which, if publicized, would cause financial harm, physical harm, or even pain or fear, are *lashon hara*… One may not live in such people's neighbourhoods, let alone sit with them and listen to them.

22. Talmud, Bava Batra 60a

מתני' לא יפתח אדם לחצר השותפין פתח כנגד פתח וחלון כנגד חלון היה קטן לא יעשנו גדול אחד לא יעשנו שנים אבל פותח הוא לרה"ר פתח כנגד פתח וחלון כנגד חלון היה קטן עושה אותו גדול ואחד עושה אותו שנים

גמ' מנהני מילי א"ר יוחנן דאמר קרא +במדבר כ"ד+ וישא בלעם את עיניו וירא את ישראל שוכן לשבטיו מה ראה ראה שאין פתחי אהליהם מכוונין זה לזה אמר ראוין הללו שתשרה עליהם שכינה ...

Mishnah: One may not open a window opposite a window, or a door opposite a door, across a yard owned in partnership. One may not increase their size, or turn one into two. However, one may do this into a public domain.

Gemara: How do we know this? Rabbi Yochanan said: "And Bilaam raised his eyes and saw Israel dwelling in its tribes." What did he see? That the entrances of their tents were not aligned. He said, "These people are suited for the Divine presence."…

<u>Case 4: Conflict of Interest</u>

23. Baker, Dellaert, *Regulating Robo Advice Across the Financial Services Industry*, Faculty Scholarship at Penn Law (2018)

While robo advisors have the potential to outperform humans in matching consumers to mass market financial products, they are not inherently immune from the misalignment of incentives that has historically affected financial product intermediaries. A robo advisor can be designed to ignore those incentives, but many consumer financial product intermediaries that develop or purchase robo advisors are subject to those incentives. It would be naïve to simply assume that intermediaries will always choose the algorithms and choice architecture that are best for consumers, rather than those that are best for the intermediaries.

24. Rabbi Yosef Karo (16th century Turkey/Israel), Shulchan Aruch, Choshen Mishpat 227:21

האומר לחבירו "על מנת שאין לך עלי אונאה," יש לו עליו אונאה. בד"א, בסתם, שאינו יודע כמה אונאה יש בו כדי שימחול... אבל במפרש, אין לו אונאה. כיצד? מוכר שאמר ללוקח, "חפץ זה שאני נותן לך בק"ק יודע אני שאינו שוה אלא מאה, על מנת שאין לך עלי אונאה אני מוכר לך," אין לו עליו אונאה...

If one says to another, "On condition that you cannot claim *ona'ah*," he can still claim *ona'ah*. This is true where he did not specify, for the other does not know how much *ona'ah* is present, to forgive it… But where he specifies, there is no claim of *ona'ah*. For example: Where a merchant tells a customer, "I am giving you this item for 200. I know it is only worth 100, but I am selling it to you on condition that you cannot claim *ona'ah* from me," one cannot claim *ona'ah*.