MACHINE LEARNING, MARKET MANIPULATION, AND COLLUSION ON CAPITAL MARKETS: WHY THE "BLACK BOX" MATTERS

ALESSIO AZZUTTI* — WOLF-GEORG RINGE** — H. SIEGFRIED STIEHL***

ABSTRACT

This Article offers a novel perspective on the implications of increasingly autonomous and "black box" algorithms, within the ramification of algorithmic trading, for the integrity of capital markets. Artificial intelligence (AI) and particularly its subfield of machine learning (ML) methods have gained immense popularity among the great public and achieved tremendous success in many real-life applications by leading to vast efficiency gains. In the financial trading domain, ML can augment human capabilities in price prediction, dynamic portfolio optimization, and other financial decision-making tasks. However, thanks to constant

---

* PhD Candidate in Law; Research Associate at the Institute of Law & Economics, University of Hamburg.

** Professor of Law and Finance; Director of the Institute of Law & Economics, University of Hamburg; Visiting Professor, University of Oxford.

*** Senior Professor at the Department of Informatics, University of Hamburg.

We are grateful for comments on earlier drafts of this paper by Pedro Batista, Maria Grigoropoulou, Christopher Ruof, and Roee Sarel, as well as participants at the Singapore Management University School of Law's workshop on AI and Fintech, the Chinese University of Hong Kong Faculty of Law's Conference on FinTech, the University of Hamburg/National Taiwan University Conference on Law and Technology, the Italian Society of Law & Economics' 15th Annual Conference, the European Banking Institute's 2020 Global Annual Conference on financial regulation, the Wharton School of the University of Pennsylvania's Conference on financial regulation, the European Centre of Excellence on the Regulation of Robotics & AI's 2021 Conference on AI regulation, and the Asia-Pacific and Europe Law Institutes Alliance's 2021 Conference on AI and international law. Funding from the Hamburg "Law, Finance, and Technology" project, sponsored by Joachim Herz Foundation, is gratefully acknowledged (https://LFT.ile-hamburg.de).

progress in the ML technology, the prospect of increasingly capable and autonomous agents to delegate operational tasks and even decision-making is now beyond mere imagination, thus opening up the possibility for approximating (truly) autonomous trading agents anytime soon.

Given these spectacular developments, this Article argues that such autonomous algorithmic traders may involve significant risks to market integrity, independent from their human experts, thanks to self-learning capabilities offered by state-of-the-art and innovative ML methods. Using the proprietary trading industry as a case study, we explore emerging threats to the application of established market abuse laws in the event of algorithmic market abuse, by taking an interdisciplinary stance between financial regulation, law and economics, and computational finance. Specifically, our analysis focuses on two emerging market abuse risks by autonomous algorithms: market manipulation and "tacit" collusion. We explore their likelihood to arise in global capital markets and evaluate related social harm as forms of market failures.

With these new risks in mind, this Article questions the adequacy of existing regulatory frameworks and enforcement mechanisms, as well as current legal rules on the governance of algorithmic trading, to cope with increasingly autonomous and ubiquitous algorithmic trading systems. We demonstrate how the "black box" nature of specific ML-powered algorithmic trading strategies can subvert existing market abuse laws, which are based upon traditional liability concepts and tests (such as "intent" and "causation"). We conclude by addressing the shortcomings of the present legal framework and develop a number of guiding principles to assist legal and policy reform in the spirit of promoting and safeguarding market integrity and safety.

Keywords: artificial intelligence, algorithmic trading, market manipulation, collusion, black box

JEL Codes: G18, G28, G38, K22, O33, O38

TABLE OF CONTENTS

I. INTRODUCTION

Thanks to artificial intelligence (AI)'s continuous progress, the sub-field of machine learning (ML) today enables the creation of increasingly "autonomous AI agents"[1] in many domains. In finance, algorithmic trading systems (ATSs) have already reached a level of enormous technological sophistication and a high degree of system automation.[2] AI, and ML methods in particular, allow for ATSs with increased autonomy to be established.[3] While having the capacity to revolutionize trading as we know it, delegating financial decision-making to increasingly autonomous and "black box" AI trading agents can also expose markets to new sources of risk.[4]

Specifically, this Article explores emerging threats to the safe application of established legal concepts of liability for market abuse in dealing with misconducts by increasingly autonomous AI trading agents, using the proprietary trading industry as a case study. As we will see, autonomous AI trading could achieve unprecedented versatility and develop unexpected capabilities beyond what human experts can reasonably expect. Indeed, thanks to self-learning, AI traders could behave in unpredictable ways, for both good and evil. As discussed below, these risks include new forms

---

[1] For "autonomous AI agents," we generally refer to agent systems in automation technology; this envisions a delimitable (hardware and/or software) unit with defined goals that the agent strives to achieve through autonomous behavior and interactions with the environment and other agents. *Cf., e.g.*, Stan Franklin & Art Graesser, *Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents*, *in* Intelligent Agents III: AGENT THEORIES, ARCHITECTURES, AND LANGUAGES 21, 25 (Jörg P. Müller, Michael J. Wooldridge & Nicholas R. Jennings eds., 1996) ("An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.").

[2] Michael P. Wellman & Uday Rajan, *Ethical Issues for Autonomous Trading Agents*, 27 MINDS & MACHINES 609, 610-11 (2017).

[3] *See* FIN. STABILITY BD., ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN FINANCIAL SERVICES: MARKET DEVELOPMENTS AND FINANCIAL STABILITY IMPLICATIONS 8 (2017) [hereinafter FSB], https://www.fsb.org/wp-content/uploads/P011117.pdf.

[4] *Id.* at 24-34 (identifying both micro and macro sources of financial stability risks led by the widespread adoption of AI in the financial services industry). *See also* Adriano Koshiyama, Nick Firoozye & Philip Treleaven, *Algorithms in Future Capital Markets: A Survey on AI, ML and Associated Algorithms in Capital Markets*, *in* PROCEEDINGS OF ACM ICAIF '20 (2020), https://dl.acm.org/doi/pdf/10.1145/3383455.3422539 (reviewing the strengths and weaknesses of certain ML algorithms applied to financial trading and discussing their future impact on global capital markets).

of market manipulation and algorithmic "tacit" collusion.[5] Notably, several ethical and legal questions arise when dealing with issues of liability for algorithms' misbehavior.[6] Our findings suggest that AI's misconduct can ultimately subvert existing prohibitions of market abuse. This study contributes to enhancing our understanding of the risks associated with liability for autonomous AI decision-making. It thus enriches the scientific debate on AI and finance, to ultimately inform global regulators when thinking about innovative regulatory solutions, taking into account the technology's specificities. There is indeed a need for a regulatory paradigm shift favoring increased adaptability *vis-à-vis* the challenges posed by a continually evolving technological market ecosystem,[7] to effectively safeguard capital markets' integrity and global financial stability.

We proceed as follows. Section II introduces the concept of autonomous "AI traders" and investigates the technological potential of their emergence. Subsequently, Section III shows that such self-learning agents may also learn how to game the system and engage in manipulative and collusive practices. Section IV illustrates how the present legal framework falls short of providing a sound response to algorithmic market manipulation and develops a number of guiding principles for reform. Section V concludes.

II. AUTONOMOUS "AI TRADERS": THE (PRESENT AND) FUTURE OF ML-POWERED ALGORITHMIC TRADING

When exploring the implications of AI trading, it is first helpful to review the current state of modern AI applications in the financial

---

5 *See infra* Section III.

6 *See generally* Thomas C. King, Nikita Aggarwal, Mariarosaria Taddeo & Luciano Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, 26 SCI. & ENG'G ETHICS 89 (2020) (offering a systematic and interdisciplinary literature review of the foreseeable threats of AI-related crimes and related ethical and legal questions); Wellman & Rajan, *supra* note 2, at 611-13 (addressing the ethical and legal issues relating to AI trading agents' misconduct).

7 *See, e.g.*, ANDREW W. LO, ADAPTIVE MARKETS: FINANCIAL EVOLUTION AT THE SPEED OF THOUGHT 365-71 (ed. 2019) (arguing that "adaptive regulation" looks at the financial system as an organic ecosystem, and rather than regulating market behaviors via traditional "command-and-control" approach, it requires to develop a better understanding of why misconducts arise and determine what aspects of the environment need to be changed to constrain them).

trading domain. In this Section, we will show that recent progress in computational finance indeed brings us closer to the development of truly autonomous algorithmic agents, based on AI, that can act independently on the capital market and learn from the outcomes of their own decisions, when given a pre-defined objective. This prepares the ground for Section III, which addresses the role of AI in facilitating new forms of algorithmic market abuse, irrespective of any direct human involvement in gaming market rules.

When considering both determinants and path-dependencies of the "algorithmic revolution" that, only in the last few decades, has shaped global capital markets' architecture and functioning,[8] there are good reasons to believe that algorithms will continue to gain an increasingly pervasive role. Indeed, the financial industry is currently undergoing profound digital transformation underpinned by AI.[9] In global finance, algorithms contribute to conducting, managing, and monitoring trading activities. Sometimes, they also cause disruptions to the safe and orderly functioning of markets.[10] Nevertheless, financial technology innovation—such as algorithmic trading—has been generally supported by regulation to foster competition among market participants on different levels, with the

---

8 *See* Tom C.W. Lin, *The New Investor*, 60 UCLA L. REV. 678 (2013) (coining the term "cyborg finance" to describe the beginning of a new era in global finance, increasingly dominated by algorithms); *see also* Marc Lenglet, *Conflicting Codes and Codings: How Algorithmic Trading is Reshaping Financial Regulation*, 28 THEORY, CULTURE & SOC'Y 44 (2011) (arguing that the algorithmization of trading on capital markets has had significant consequences for the nature of financial regulation).

9 *See* FSB, *supra* note 3; CARSTEN JUNG, HENRIKE MUELLER, SIMONE PEDEMONTE, SIMONE PLANCES & OLIVER THREW, BANK OF ENG. & U.K. FIN. CONDUCT AUTH. (FCA), MACHINE LEARNING IN UK FINANCIAL SERVICES (2020) [hereinafter FCA], https://www.bankofengland.co.uk/-/media/boe/files/report/2019/machine-learning-in-uk-financial-services.pdf?la=en&hash=F8CA6EE7A5A9E0CB182F5D568E033F0EB2D21246 [https://perma.cc/W24Z-DE87] (surveying more than 100 financial firms within the UK financial system); CAMBRIDGE CTR. FOR ALT. FIN. & WORLD ECON. FORUM, TRANSFORMING PARADIGMS: A GLOBAL AI IN FINANCIAL SERVICES SURVEY (2020), http://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf [https://perma.cc/S3GT-B2PD] (surveying around 150 financial firms from thirty-three different countries, and reporting that seventy-seven percent of all respondents believe AI will have paramount importance for their business models within the next two years).

10 *See generally* Andrei A. Kirilenko & Andrew W. Lo, *Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents*, 27 J. ECON. PERSPS. 51 (2013) (providing a brief survey of algorithmic trading, from its origins to fist accidents caused to markets); Yesha Yadav, *How Algorithmic Trading Undermines Efficiency in Capital Markets*, 68 VAND. L. REV. 1607 (2015) (describing how algorithmic trading has radically transformed markets and hampered prices' informativeness and their ability to serve allocative efficiency).

objective of supporting the development of more efficient and liquid markets.[11]

The proliferation of markets and financial assets, as well as the acceleration of trading speed, are all fundamental factors contributing to generating a massive amount of granular and high-frequency data. Notably, to find profitable investment opportunities, AI trading can exploit massive data that are no more intelligible for the human mind.[12] Useful data for AI today come in very different forms and levels of quality (beyond traditional financial data, such as fundamental data or market data and their derivatives), with "alternative data"[13] taking on growing importance.[14] Originally, algorithmic trading was based on deterministic "rule-based" systems, which are notoriously constrained by human experts' knowledge and assumptions, both tacit and explicit, about specific domains.[15] Thanks to simultaneous progress made in high-performance computing and communication (e.g., edge/cloud computing) and Big Data analytics, ML methods today allow for trading algorithms to be far more flexible to changing market conditions, under different levels of autonomy.[16] ML and Big Data are together the fundamental ingredients of most innovative and cutting-edge algorithmic trading strategies.[17]

---

11 On the role of financial regulation to foster innovation while safeguarding competition and other public goals, see generally Wolf-Georg Ringe & Christopher Ruof, *Regulating Fintech in the EU: The Case for a Guided Sandbox*, 11 EUR. J. RISK REGUL. 604 (2020).

12 FSB, *supra* note 3, at 18 (reporting on the use of AI and machine learning to devise trading and portfolio management strategies).

13 *See* MARKO KOLANOVIC & RAJESH T. KRISHNAMACHARI, J.P. MORGAN, BIG DATA AND AI STRATEGIES: MACHINE LEARNING AND ALTERNATIVE DATA APPROACH TO INVESTING 28-50 (2017) https://www.cfasociety.org/cleveland/Lists/Events%20Calendar/Attachments/1045/BIG-Data_AI-JPMmay2017.pdf (providing a comprehensive overview of different kinds of alternative data, their taxonomy, and possible use for specific trading strategies).

14 *See* MARCOS LÓPEZ DE PRADO, ADVANCES IN FINANCIAL MACHINE LEARNING 23-25 (2018).

15 For a comprehensive overview on ATSs, and their different components and operational functioning, see generally Philip Treleaven, Michal Galas & Vidhi Lalchand, *Algorithmic Trading Review*, 56 COMMC'NS ACM 76 (2013).

16 *See* FCA, *supra* note 9, at 2; *see also* Kolanovic & Krishnamachari, *supra* note 13, at 9-11.

17 These include strategies such as: (i) *signal processing*, the art of filtering meaningful information from noisy data to discern trading patterns; (ii) *market sentiment* analysis, a strategy that extrapolates markets appetite for trading by

However, the prospect of fully autonomous AI agents is still assumed today to be beyond imagination.

a. Towards Autonomous Trading Agents

ML can assist investment firms in both pattern recognition and financial decision-making tasks. According to key differences in the fields of algorithms learning from data, which also relates to the varying degree of human involvement, three basic ML paradigms exist. First, in "supervised learning" (SL) methods, which can be used for regression and classification purposes, users need to train their algorithms with pre-labeled empirical data, meaning that the correct outputs for all trading data are known in advance. Once a general rule has been learned, it has to be carefully validated and tested before it is applied to, as an example, predictive trading tasks.[18] For instance, under SL, algorithms can use technical market indicators or other useful data to predict the next day's winning and losing stocks from past observations yielded from empirical data.[19] Secondly, in "unsupervised learning" (UL), which is instead used for clustering and factor analyses, algorithms work without any pre-labeled data provided by a human expert.[20] Under this ML method, algorithms autonomously infer patterns (e.g., "regularity") in the data with similar distinctive features.[21] An ATS can jointly integrate both SL and UL to solve different trading tasks. For instance, UL algorithms can preliminarily perform a cluster analysis to extract features from data to identify trading opportunities. The result is

learning from market activity; (iii) *news reader*, which leverages on the role of news from different media to look for investment opportunities; and (iv) *pattern recognition*, or the computational ability to learn from changing price patterns on markets how to classify different market prices dynamics in order to anticipate price movements to gain a profit. BONNIE G. BUCHANAN, ALAN TURING INST., ARTIFICIAL INTELLIGENCE IN FINANCE 16 (2019), https://www.turing.ac.uk/sites/default/files/2019-04/artificial_intelligence_in_finance_-_turing_report_0.pdf [https://perma.cc/2HC3-WM9X].

18 *See* Kolanovic & Krishnamachari, *supra* note 13, at 18.

19 *See* Kolanovic & Krishnamachari, *supra* note 13, at 57 and 77 (discussing, in technical detail, the functioning of supervised learning methods for regression and classification purposes).

20 *See* Kolanovic & Krishnamachari, *supra* note 13, at 18.

21 *See* Kolanovic & Krishnamachari, *supra* note 13, at 93-101 (discussing, in technical detail, the functioning of unsupervised learning methods for clustering and factor analyses purposes).

then passed, as input data, to the supervised learning component for further computational steps, like stock price prediction.[22] Thereafter, the AI system is ready to execute trading. Thus, both ML methods can assist investment firms in automating trading in financial instruments. However, neither yet achieves autonomy in ATSs, since some human assistance is still usually required to face evolving market conditions, such as tail risk and unobserved market events.[23] In fact, both methods are simply constrained by the empirical nature of data. In contrast, although humans can infer actions from their past experiences, they are known to also rely on, for instance, hardly explicable intuition and gut feeling for decision-making under conditions of uncertainty.[24]

---

[22] For an early study combining supervised and unsupervised learning in a hybrid strategy, see Cheng-Lung Huang & Cheng-Yi Tsai, *A Hybrid SOFM-SVR with a Filter-Based Feature Selection for Stock Market Forecasting*, 36 EXPERT SYS. WITH APPLICATIONS 1529 (2009) (combining an unsupervised learning algorithmic component, responsible for filter-based feature selection to choose important input attributes, with a supervised learning one that is subsequently called upon to predict stock market prices index-based).

[23] *See* John Moody, Lizhong Wu, Yuansong Liao & Matthew Saffell, *Performance Functions and Reinforcement Learning for Trading Systems and Portfolios,* 17 J. FORECASTING 441, 442 (1998) (highlighting the fundamental trading policy misalignment between supervised learning methods' optimisation goal, which is constrained to what the algorithm can observe and learn from historical data, with the ultimate objective of a general investor that instead faces changing time-dependent constraints due to constantly evolving market dynamics); *see also* Quang-Vinh Dang, *Reinforcement Learning in Stock Trading*, *in* ADVANCED COMPUTATIONAL METHODS FOR KNOWLEDGE ENGINEERING 311, 312 (Hoai An Le Thi, Hoai Minh Le, Tao Pham Dinh & Ngoc Thanh Nguyen eds., 2019) (highlighting the inadequacy to deal with time-delayed rewards as the main technical limitation of supervised learning methods, these being constrained to only achieving the best prediction at an exact local point in time and without considering any delayed reward or punishment, as a result of which, supervised learning methods applied to financial decision-making can only provide actionable recommendations, rather than an entirely and effectively autonomous ATS).

[24] For a behavioral economics study on the influence of emotions on the decision making and performance of professional traders, see generally Mark Fenton-O'Creevy, Emma Soane, Nigel Nicholson & Paul Willman, *Thinking, Feeling and Deciding: The Influence on the Decision Making and Performance of Traders*, 32 J. ORG. BEHAV. 1044 (2011), finding that experienced traders have a relatively high meta-cognitive engagement with emotion regulation, allowing them to discriminate between emotions in terms of their relevance to the decision at hand and how to deal with them to enhance performance effectively. *But see* Andrew W. Lo, Dmitry V. Repin & Brett N. Steenbarger, *Fear and Greed in Financial Markets: A Clinical Study of Day-Traders*, 95 AM. ECON. REV. 352 (2005) (suggesting that emotions can instead negatively impact on trading performance and that, in contrast, successful trading may be due to a reduced level of emotional reactivity).

Finally, and most importantly, a third ML paradigm under the name of "reinforcement learning" (RL) has emerged to overcome some of these limitations.[25] RL is the most advanced of the ML paradigms in the context of our analysis below, as it lies at the foundation of autonomous (software) agents. This very heterogenic ML category encompasses computational approaches that allow algorithms to learn, through a "trial-and-error" process, within an uncertain and dynamic environment. In doing so, RL agents are called to take action with the ultimate goal to realize a pre-defined objective or optimize a cost or utility function pursuant to that objective. In addition, as is the case in a real market context, RL agents need to take into account the implications of their own behaviors. In other words, they are goal-oriented and face a constant trade-off between "exploration" and "exploitation" in the space and/or time of a particular domain. Thus, RL agents must "exploit" actions that were learned in the past to achieve the best rewards. At the same time, exploiting implies the ability to "explore" in advance the best policies among all options, both known and unknown, in order to make better decisions in the future.[26] In a financial trading context, RL allows the "forecasting" and "portfolio construction" tasks to be integrated, thus aligning the ML problem with the investors' ultimate goal.[27] In fact, unlike (un)supervised methods, in which ML is used for generalization purposes, RL agents aim to learn best policy actions that maximize the likelihood of a long-term goal being achieved while also taking into account real markets' constraints, such as liquidity and transaction costs.[28] In a manner of speaking, RL attempts to resemble how human traders traditionally act on financial markets and learn from their own trading experiences and strategies to pursue their profit-maximizing objectives. The computational finance literature has developed several RL applications for trading, categorized according to the exact optimizing method employed in

---

[25] *See generally* RICHARD S. SUTTON & ANDREW G. BARTO, REINFORCEMENT LEARNING: AN INTRODUCTION (2d ed. 2017) (providing a thoughtful technical introduction to RL methods).

[26] *Id.* at 1-5.

[27] Thomas G. Fischer, *Reinforcement Learning in Financial Markets – A Survey* 2 (Friedrich-Alexander-Universität Erlangen-Nürnberg Instit. for Econ., Working Paper No. 12, 2018), https://www.econstor.eu/bitstream/10419/183139/1/1032172355.pdf [https://perma.cc/WL44-EPBJ].

[28] *Id*.

the self-learning process.[29] Not surprisingly, therefore, RL has already had an enormous impact on optimizing financial trading tasks, with promising results in high-frequency trading (HFT).[30]

Lately, much of the hype surrounding AI has been about "deep learning" methods, a more recent sub-field in ML.[31] Deep learning is based on "artificial neural networks" (ANNs)—i.e., mathematical models that by and large resemble the neuronal structure and functioning of the human cortex—which aim to best approximate input data by learning on multiple abstraction levels (cf. "convolutional neural network" methods).[32] ANNs can be used in combination with SL and RL methods and are proposed to achieve greater accuracy and predictive power in our application domain,[33] albeit like other ML methods they can nevertheless be exposed to human bias.[34] However, there can be other drawbacks, since,

---

[29] However, all these different methods share the same core idea: to develop a mathematical model that can plan future actions while also considering whether and how the own actions will impact the market. Nevertheless, in developing RL methods, the real challenge is to find meaningful data for the above formalisation. *Cf. id.* at 3-35 (providing a detailed discussion on the three RL main paradigms, namely the "critic," "actor-only," and "actor-critic," and explaining how these different RL methods deal with the mathematical problems of modeling the three core RL models' components such as "state," "action," and "space").

[30] *See* Michael Kearns & Yuriy Nevmyvaka, *Machine Learning for Market Microstructure and High Frequency Trading*, *in* HIGH FREQUENCY TRADING: NEW REALITIES FOR TRADERS, MARKETS AND REGULATORS 91, 92 (David Easley, Marcos Lopez de Prado & Maureen O'Hara eds., 2013) (illustrating three RL applications to HFT problems, such as (i) optimized trade execution; (ii) predicting price movements from order book state; and (iii) optimized execution in dark pools via censorship exploration).

[31] William Magnuson, *Artificial Financial Intelligence*, 10 HARV. BUS. L. REV. 337, 344 (2020). For an introduction to deep learning methods and their technicalities, see generally Yann LeCun, Yoshua Bengio & Geoffrey Hinton, *Deep Learning*, 521 NATURE 436 (2015) (discussing various methods of deep learning, including supervised learning, backpropagation, convolutional neural network, and distributed representation and language processing).

[32] *Cf.* Li Deng & Dong Yu, *Deep Learning: Methods and Applications*, 7 FOUNDS. & TRENDS SIGNAL PROCESSING 197, 224 (2013).

[33] *See generally* Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek & Omer Berat Sezer, *Deep Learning for Financial Applications: A Survey*, 93 APPLIED SOFT COMPUTING J. 1, 1 (2020) (providing an exhaustive review of deep learning methods combined with supervised learning and reinforcement learning, with many algorithmic trading examples).

[34] This is notwithstanding the effects of the so-called "inductive bias," i.e., the series of assumptions made by the model to learn the target function and generalize from training data. *See* Anirudh Goyal & Yoshua Bengio, Inductive Biases for Deep Learning of Higher-Level Cognition 3 (Feb. 17, 2021) (unpublished manuscript) (on file with authors) (examining the role of different inductive biases that can be used

besides a greater propensity towards "overfitting,"[35] these ML methods are accompanied by the so-called problem of the "black box."[36] The "black box" problem is where both the developers and users of AI may not fully understand and explain *why* and *how* their algorithms have generated a particular output given specific data input.[37] The "black box" problem is often framed in terms of issues of AI "transparency," "explainability," and "trustworthiness," especially for ML-based decision-making in critical domains related to human life,[38] which underpins the problem of "auditability" and "accountability" in cases of AI wrongdoing.[39] As we will see in Section IV, the "black box" problem is central to our assessment of existing legal systems' ability to effectively cope with circumstances of market abuse by autonomous AI traders.

The combination of "deep" *and* "reinforcement learning" techniques allows for the creation of so-called "deep reinforcement

---

to encourage the learning process of deep learning methods to prioritize solutions according to certain properties).

[35] "Overfitting" refers to the problem that models are too specific to training data that can generalize poorly on new datasets; as such, overfitting models cannot safely be applied on real markets. *See* Shihao Gu, Bryan Kelly & Dacheng Xiu, *Empirical Asset Pricing via Machine Learning*, 33 REV. FIN. STUD. 2223, 2225 (2020); *see also* LÓPEZ DE PRADO, *supra* note 14, at 151-56 (highlighting the fundamental role of backtesting techniques to prevent overfitting, as well as the ability by users to understand the importance of data features for their models).

[36] *See generally* Zachary C. Lipton, *The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery*, ACM QUEUE, May-June 2018 (discussing the many possible facets of the concept of AI "black box" model "interpretability"). *But see* LÓPEZ DE PRADO, *supra* note 14, at 15-16, 114 (dismissing the "black box" argument as a misconception, and arguing that any conscious use of ML in financial trading is only possible by means of "white boxes" algorithms, which also requires a sound understanding on data feature and their importance for the model).

[37] For a concise explanation of the "black box" problem in AI decision-making, see generally Dino Pedreschi et al., *Meaningful Explanations of Black Box AI Decision Systems*, THIRTY-THIRD AAAI CONFERENCE ON A.I. (AAAI-19) 9780 (2019) (discussing the "black box" problem from an ethical perspective, and reviewing both technical challenges and possible solutions to achieve meaningful explainability in opaque ML-based systems).

[38] *See generally* Eur. Comm'n, HIGH-LEVEL GROUP ON ARTIFICIAL INTELLIGENCE: ETHICS GUIDELINES FOR TRUSTWORTHY AI (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 [perma.cc/J6AU-A789].

[39] *See* Alejandro Barredo Arrieta et al., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI*, 58 INFO. FUSION 82, 84 (2020) (offering an overview of recent efforts to achieve AI explainability and stressing the relevance of AI explainability to guarantee effective auditability and accountability in relation to different AI stakeholders).

learning" (DRL) methods. By combining the upsides of these two ML paradigms, DRL algorithms are able to take in very large datasets, find latent correlations thanks to deep learning, and learn to decide which actions to perform in order to optimize a function via RL in pursuit of a pre-defined objective.[40] Autonomous agents based on DRL have achieved tremendous success by showing superior-to-human capabilities in many real-life settings, including video[41] and board games,[42] among others.[43] With that in mind, DRL methods could arguably be used to achieve autonomous AI trading agents, eventually implying the exclusion of human control as the last resort. Within the scientific community, a growing amount of published work has been applying DRL agents to financial trading problems.[44] Under DRL, for instance, AI traders can, first, gain data-driven insights about a complex and dynamic trading environment via DL and, second, use RL to flexibly learn optimal trading strategies solely through their trading activities on markets, which provide constant feedback on their performance.[45] But the possibilities do not end here: in principle, several ML components can be integrated into DRL-based ensemble strategies to achieve

---

40 For a first introduction to DRL algorithms, see Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage & Anil Anthony Bharath, *A Brief Survey of Deep Reinforcement Learning*, IEEE SIGNAL PROCESSING MAG., Sept. 27, 2017, at 1-13.

41 *See* Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 NATURE 529, 529-530 (2015) (developing a deep Q-network algorithmic agent that, fed with games' graphic pixels and score as the only input data, achieved professional human gamers' ability across a set of forty-nine different Atari 2600 games).

42 *See* David Silver et al., *Mastering the Game of Go with Deep Neural Networks and Tree Search*, 529 NATURE 484, 484 (2016) (developing the first DRL agents able to defeat a human professional player in the full-sized game of Go by training them through a novel combination of supervised learning from human expert games and reinforcement learning from simulated games).

43 For an early overview of different DRL methods and their successful application in different real-life settings, see generally Arulkumaran et al., *supra* note 40 (providing an overview of the field of reinforcement learning including deep-Q network, trust region policy optimization, and asynchronous advantage actor-critic).

44 *See* Zihao Zhang, Stefan Zohren & Stephen Roberts, *Deep Reinforcement Learning for Trading*, J. FIN. DATA SCI. 25 (2020) (designing a deep reinforcement learning algorithmic agent to derive trading strategies for continuous future contracts).

45 *See generally* Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren & Qionghai Dai, *Deep Direct Reinforcement Learning for Financial Signal Representation and Trading*, 28 IEEE TRANSACTIONS ON NEURAL NETWORK & LEARNING SYS. 653 (2017) (discussing how to train AI to trade through a recurrent deep neural network for real time financial signal representation and trading).

different levels of system sophistication and autonomy.[46] Various AI agents can be combined in multi-agent systems to benefit from their different skill specializations,[47] or in ensemble strategies where they need to act jointly to achieve best performance.[48] It follows that current research in computational finance provides initial evidence about DRL methods as main ML frameworks for the successful implementation of increasingly capable and autonomous AI trading agents.

b. Ongoing Progress and Challenges

Given all this progress made in theories, methods, and technologies, it is worth emphasizing that algorithmic trading agents are called upon to operate within a complex and dynamic market environment. Real markets can be substantially different and more complex than in-lab simulation environments, making it hard to effectively and safely develop autonomous AI agents for real-life applications. In fact, the successful implementation of AI agents via RL methods requires taking into account several limitations, among which the "curse of dimensionality"[49] is only one

---

[46] *See* Ozbayoglu et al., *supra* note 33, at 6.

[47] *See generally* Salvatore Carta, Andrea Corriga, Anselmo Ferreira, Alessandro Sebastian Podda & Diego Reforgiato Recupero, *A Multi-Layer and Multi-Ensemble Stock Trader Using Deep Learning and Deep Reinforcement Learning*, 51 APPLIED INTEL. 889, 889-905 (2020) (developing a multi-layer and multi-ensemble stock trading agent by combining deep learning and reinforcement learning methods in a unique strategy to trade on futures markets).

[48] *See generally* Salvatore Carta, Anselmo Ferreira, Alessandro Sebastian Podda, Diego Reforgiato Recupero & Antonio Sanna, *Multi-DQN: An Ensemble of Deep Q-learning Agents for Stock Market Forecasting*, 164 EXPERT SYS. WITH APPLICATIONS (2021) (proposing a novel ensemble of same DRL algorithms that, by developing different experiences on market dynamics, engage in cooperative tasks to reach best policy actions by agreement on competing strategies).

[49] In reinforcement learning methods, the "curse of dimensionality" refers to the problem arising when RL agents are called to learn from a too large or continuous "environment," i.e., in a high-dimensional data space. *See, e.g.*, Vangelis Bacoyannis, Vacslav Glukhov, Tom Jin, Jonathan Kochems & Doo Re Song, Idiosyncrasies and Challenges of Data Driven Learning in Electronic Trading 4-5 (NIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services) (unpublished manuscript) (on file with authors) ("Describing the market state of the limit order book is a variable dimension and high dimension problem."). *But see, e.g.*, Silvio Barra, Salvatore Mario Carta, Andrea Corriga, Alessandro Sebastian Podda & Diego Reforgiato Recupero, *Deep Learning and Time Series-to-Image Encoding for Financial Forecasting*, 7 IEEE/CAA J. AUTOMATICA SINICA 683

facet.[50] More generally, there are also fundamental challenges in assessing the quality of ML research applied to financial trading. While ML research is successfully expanding, computational finance literature has so far failed to provide a convincing scientific framework or even methodology to analyze different ML methods (i.e., theoretical limits, accuracy, and experimental success and failure).[51] Unlike other AI fields of application, no clear benchmark exists yet to assess and compare competing ML algorithms for financial trading.[52] Apart from this, proprietary details regarding the nature and role of the utilized empirical data as well as information about the learning process itself (or "hyper parameters" in general) further complicate or even prohibit the comparison of different ML research findings,[53] thus rendering the replication of ML results impossible.

Moreover, AI traders' autonomy and complexity also exacerbate the agency problem in algorithmic trading. Financial laws usually require trading algorithms to produce predictable, controllable, and explainable trading behavior, not least to avoid disrupting financial markets' orderly functioning.[54] Accordingly, users of AI should be able to explain how AI systems reach their optimized trading strategies to comply with financial law and regulation, including taking accountability for affecting clients, consumers, or the public.

Despite all these difficulties, we believe that it is realistic to expect autonomous AI traders to become a reality on trading floors one day. Once they become a reality, a number of the policy issues mentioned above will come to the fore. To start, acknowledging that the most innovative ML research advancements are likely to emerge within investment firms' proprietary projects, protected by

(2020) (addressing the issue by employing "Gramian angular fields"—GAF—images generated from time series, a novel computational approach intended to simplify computational processes).

50 *See* Fischer, *supra* note 27, at 38-39 (providing an overview of both strengths and weaknesses for different reinforcement learning methods).

51 For an overview of the most relevant limitations in the ML research applied to financial forecasting, see generally Spyros Makridakis, Evangelos Spiliotis & Vassilios Assimakopoulos, *Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward*, PLOS ONE, Mar. 27, 2018.

52 *See* Lukas Ryll & Sebastian Seidens, *Evaluating the Performance of Machine Learning Algorithms in Financial Market Forecasting: A Comprehensive Survey* 1-2 (July 6, 2019) (unpublished manuscript) (on file with authors).

53 *Id.*

54 *Cf.* Bacoyannis et al., *supra* note 49, at 5-6 (arguing that "hierarchical" reinforcement learning methods could assist investment firms in compliance tasks).

intellectual property rights, is certainly not reassuring from a policy perspective. Indeed, and given that academic research is openly accessible but still limited in scope, there are significant risks that the AI technology in financial trading may evolve without any sound normative considerations or even academic and public scrutiny. While AI is undoubtedly proposed as a game-changer for trading on capital markets, both regulators and market participants could become concerned about specific ML methods and applications leading to greater uncertainties and novel risks. Indeed, this is the first time in human history that we are delegating cognitive agency to algorithms to be utilized in critical domains despite knowing that, in the worst-case scenario, we could become incapable of controlling their functioning.

III. ALGORITHMIC MARKET ABUSE BY AUTONOMOUS AI TRADERS

We have seen above that the technological potential for the emergence of truly autonomous agents that trade on capital markets is realistic. Furthermore, we have shown that the most advanced of these trading machines will be able to learn and refine a particular investment strategy independently, given a pre-defined goal (most likely, profit maximization). This seemingly positive development has a dark side, however, where investment decisions by independent algorithms could be used for trading strategies that undermine the laws of capital markets—and would be applied to maximize profit from manipulative practices or collusion.

With the rise of algorithmic trading, innovative manipulative schemes inevitably arise, and forms of algorithmic manipulation have indeed emerged already.[55] With the prospect of fully autonomous AI traders proliferating global capital markets sometime soon, new and unprecedented algorithmic crime scenarios can also arise. Precisely with these risks in mind, this Section deals with new forms of market manipulation by autonomous AI traders, including new abusive cartel-like

---

[55] *See* Tom C.W. Lin, *The New Market Manipulation*, 66 EMORY L.J. 1253, 1288 (2017) (coining the term "cybernetic market manipulation" to describe both old and new forms market manipulation strategies by means of trading algorithms).

scenarios,[56] their likelihood according to both markets' microstructure and AI technical limitations, and related social harm as a consequence of market failures.[57]

a. Old and New Algorithmic Market Manipulation

The institutional role of capital markets is—as an ideal—to allow for the efficient allocation of financial resources and appropriate risk-sharing among market participants. In contrast, due to potentially being harmful to financial markets' efficiency and detrimental to public confidence in their proper functioning, some market conduct is considered illegitimate *per se*.[58] As a form of market abuse, for instance, market manipulation represents an illegitimate expression of market conduct. More specifically, market manipulation refers to any conscious attempt to interfere with the free and fair nature of trading activity, which must characterize the ordinary functioning of capital markets. As economic phenomena, manipulative conducts aim to alter artificially the price of one or more financial instruments or to influence natural forces of market activity with deceptive means to induce other investors to trade.[59] Therefore, in the absence of a market mechanism able to deal with market abuses such as market manipulation, regulatory intervention seeks to nudge market participants towards positive

---

[56] There are growing concerns among competition law scholars on autonomous AI pricing agents' ability to lead to new forms of collusive outcomes. *See generally* Ariel Ezrachi & Maurice E. Stucke, *Artificial Intelligence & Collusion: When Computers Inhibit Competition*, 2017 U. ILL. L. REV. 1775 (2017) (exploring four basic scenarios where AI can foster anti-competitive market behavior between competing firms, including what they define as the "Digital Eye" scenario in which collusion arise from autonomous AI decision-making as a rational strategy); *see also* Salil K. Mehra, *Antitrust and the Robo-Seller: Competition in the Time of Algorithms*, 100 MINN. L. REV. 1323 (2016) (discussing how pricing algorithms have changed specific industries' market dynamics and how increased autonomy in algorithmic decision-making threatens traditional competition law concepts and enforcement tools).

[57] While few tentative studies deal with both ethical and legal issues of market manipulation by autonomous AI traders, to the best of our knowledge, this is the first study addressing the risk of algorithmic "tacit" collusion on capital markets. For other similar studies, see, for example, Wellman & Rajan, *supra* note 2, at 616 (discussing the risks of autonomous AI trading engaging in "statistical arbitrage" strategies).

[58] EMILIOS AVGOULEAS, THE MECHANICS AND REGULATION OF MARKET ABUSE: A LEGAL AND ECONOMIC ANALYSIS 148 (2005).

[59] *Id.* at 107.

behaviors.[60] Whenever this does not have the desired deterrent effect, the law punishes transgressors with sanctions. As the economic nature and socially harmful effects of market manipulation are well known,[61] legal systems generally contain statutory prohibitions from specific financial laws against these forms of market abuse,[62] with the only difference being that they provide different legal definitions and scope of application of the prohibition of market manipulation.[63]

---

[60] *But see* discussion *infra* Section IV.D.

[61] *See generally,* Merritt B. Fox, Lawrence R. Glosten & Gabriel V Rauterberg, *Stock Market Manipulation and Its Regulation*, 35 YALE J. ON REGUL. 67 (2018) (addressing the harmful welfare effect of open market manipulation, and how financial law and regulation should deal with these forms of market abuse).

[62] In the U.S. legal system, the prohibition against market manipulation is expressed in a number of statutes and legal provisions. On the one hand, Section 9(a)(2) of the Securities and Exchange Act prohibits transactions in securities that create (i) "actual or apparent active trading in such security, or raising or depressing the price of such security," (ii) "for the purpose of inducing the purchase or sale of such security by others." *See* 15 U.S.C. § 78i(a)(2). While this provision seems well suited to deal with market manipulation cases, its jurisprudence has failed to provide a clear legal framework on what really constitutes an "illegitimate purpose" to trade. *See, e.g.*, Fox et al. *supra* note 61, at 114-17 (discussing the U.S. case law on the matter and arguing that this rule has only had a minimal role in developing the manipulation jurisprudence). On the other hand, Section 10(b) has found a greater scope of application against market manipulation. Notably, Section 10(b) (and particularly Rule 10b-5) deals with the prohibition of "manipulative" conduct that operates or attempts to operate as fraud or deceit upon other market participants. *See* 15 U.S.C. § 78j(b) and 17 C.F.R. § 240.10b-5. Notwithstanding its broad application scope to different manipulative strategies, Section 10(b) focuses on fraud-like market abuse. Indeed, there is an intense split between legal scholars and courts on whether, absent an additional act or omission, trading activity alone can amount to market manipulation. See Fox et al., *supra* note 61, at 118-22 (summarizing the ongoing debate among U.S. courts on the need to prove an additional element to trading alone, in order to configure and prosecute against cases of market manipulation). Finally, the Commodity and Exchange Act contains specific legal provisions on the prohibition of market manipulation of commodity prices. For instance, the 2010 Dodd-Frank Wall Street Reform and Consumer Protection Act has codified the prohibition of "spoofing" (i.e., intentionally submitting trading orders with the intention to cancel them before execution to deceive other market participants). *See, e.g.*, Gregory Scopino, *Do Automated Trading Systems Dream of Manipulating the Price of Futures Contracts? Policing Markets for Improper Trading Practices by Algorithmic Robots*, 67 FLA. L. REV. 221, 293 (2016) (assessing the merits of the Dodd-Frank reform and discussing the general issues in dealing with intent-based tests to deal with algorithmic market manipulation).

[63] The scope of application of market abuse regulations differs across jurisdictions. For instance, in the U.S., the above-mentioned Section 9(a)(2) does not apply to over-the-counter financial instruments. In the EU, instead, spot FX contracts are not protected by MAR. *See* Regulation (EU) No. 596/2014 of the European Parliament and of the Council of 16 April 2014 on Market Abuse (Market

i. Autonomous AI-Style Market Manipulations

As computational finance technology evolves, malicious actors are presented with great incentives to refine and craft new tools to manipulate markets. That humans can employ algorithms for unlawful purposes is nothing new in finance. There is indeed growing empirical evidence,[64] as well as a number of litigation cases, concerning the liability of market participants that successfully manipulate markets through algorithmic—and, particularly, HFT—strategies.[65]

Since AI can help investment firms to optimize their business operations, delegating financial trading decision-making to AI systems can arguably lead to optimized algorithmic manipulation strategies and result in very profitable trading solutions.[66] Of course, manipulation can involve significant costs and risks before any profit can materialize.[67] Consequently, investing enormous resources to train AI trading systems to learn manipulative strategies, either from historical or simulated examples or through an online observation of market dynamics, might not be worth the

---

Abuse Regulation) and Repealing Directive 2003/6/EC of the European Parliament and of the Council and Commission Directives 2003/124/EC, 2003/125/EC and 2004/72/EC, art. 2(1) and (2), 2014 O.J. (L173) 2, 3; *see also* MAR REVIEW REPORT, EUROPEAN SECURITIES AND MARKETS AUTHORITY (ESMA) 26 (2020), https://www.esma.europa.eu/sites/default/files/library/esma70-156-2391_final_report_-_mar_review.pdf [https://perma.cc/NN8D-7C86] (comparing the EU regulatory regime applied to spot FX contracts to the Australian one, wherein FX falls within the scope of market abuse regulations).

64 *See, e.g.*, Jiading Gai, Chen Yao & Mao Ye, *The Externalities of High-Frequency Trading*, 6-7 (WBS Fin. Grp. Rsch. Paper, Working Paper No. 180, 2013) (providing interesting empirical evidence about some HFT manipulative strategies on U.S.-traded stocks).

65 For a recent case involving U.S. authorities prosecuting one of the leading global investment firms, JP Morgan Chase & Co., which was found guilty of manipulating the price of U.S. Treasury securities by trading strategies aimed at misleading other market participants, see J.P. Morgan Sec. LLC, File No. 3-20094 (2020) (admin. order).

66 *See* Jón Daníelsson, Robert Macrae & Andreas Uthemann, *Artificial Intelligence and Systemic Risk*, J. BANKING & FIN. (forthcoming 2021) (manuscript at 1-2) (on file with authors) (arguing that using AI by malicious actors to optimize their techniques to harm markets is a major concern for financial stability).

67 Daniel R. Fischel & David J. Ross, *Should the Law Prohibit "Manipulation" in Financial Markets?* 105 HARV. L. REV. 503 (1991) (arguing that the law should rule out the prohibition of trade-based market manipulation, as it is usually hard to distinguish this unlawful trading conduct from legitimate trading activity, and because any attempt to market manipulation can always result in uneconomic outcomes for perpetrators).

financial commitment, given all involved risks at stake (e.g., market, operational, legal, and reputational risks). However, thanks to continuous progress being made in the optimization capabilities of specific ML methods (i.e., deep learning), increasingly autonomous AI trading systems could lead to even trickier manipulative scenarios. Autonomous AI agents could learn and discover both old and new ways to exploit market rules while pursuing their profit-maximizing objectives as an optimal and rational strategy, irrespective of the prior intent of the human developers or users.[68] Yet, cases of prosecution for liability for algorithmic market manipulation do not shed much light on the actual degree of autonomy and sophistication of the algorithms employed by malicious actors. As intellectual propriety rights generally protect algorithms' codes, we cannot expect proprietary trading firms to disclose precious details about the inner functioning of their "black box" ML algorithms and trading techniques. Nevertheless, it seems reasonable to believe that AI can offer malicious actors a broader spectrum of opportunistic strategies with which to game markets. For only this reason, market conduct authorities should start to identify and assess new risks arising from the use of increasingly capable and autonomous AI solutions for financial trading.

ii. Case Studies

Without purporting to be an exhaustive list, in this section we consider specific algorithmic trading strategies that seem to be well-suited for AI-style market manipulation.[69] Admittedly, the

---

[68] *See* Takanobu Mizuta, *Can an AI Perform Market Manipulation at Its Own Discretion? – A Genetic Algorithm Learns in an Artificial Market Simulation*, 2020 IEEE SYMPOSIUM SERIES ON COMPUTATIONAL INTEL. 407 (2020) (developing a deep learning agent, using genetic algorithms, able to discover market manipulation as an optimal strategy in an artificial market simulation); *see also* Enrique Martínez-Miranda et al., *Learning Unfair Trading: A Market Manipulation Analysis from the Reinforcement Learning Perspective*, 2016 IEEE CONFERENCE ON EVOLVING & ADAPTIVE INTELLIGENT SYS. 103 (2016) (investigating the causes that can lead a reinforcement learning trading agent to discover and enter manipulative strategies, such as "spoofing" or "pinging").

[69] Very little indeed is known about new and emerging algorithmic forms of market manipulation. *See* Tālis J. Putniņš, *An Overview of Market Manipulation*, *in* CORRUPTION AND FRAUD IN FINANCIAL MARKETS: MALPRACTICE, MISCONDUCT AND MANIPULATION 13, 35-37 (Carol Alexander & Douglas Cumming eds., 2020) (arguing that new legal questions will surely arise for cases of market manipulation by autonomous AI trading agents).

proprietary HFT industry is a market segment where AI traders could achieve great results, both for good and evil. In fact, HFT markets produce, at frightening speed, massive market microstructure data that AI-based trading strategies could more easily exploit also for unlawful purposes. For instance, advancements in AI can bring forth new, or optimize known, "deceptive" and "aggressive" HFT strategies,[70] which have been under increasing scrutiny from regulators. Indeed, AI could optimize those strategies that seek to deceive other market participants through fast submission and cancellation of orders,[71] such as "spoofing." [72] Significantly, order-based algorithmic strategies are somehow constrained by venues' rules limiting the "orders to transactions ratio" (OTR), aimed at disadvantaging *non bona fide* trading orders. However, by algorithmic learning, autonomous AI traders could find ways to optimize manipulation strategies within OTR statutory limits. Indeed, recent in-lab market simulations involving "reinforcement learning" agents provide some first evidence. Manipulation strategies like "spoofing" can emerge via RL without any prior human intent: AI traders' exploring of markets' microstructures by placing false orders would eventually be learned and applied as a profitable and rational strategy, which could, in turn, be "exploited" to maximize profits.[73]

Elsewhere, autonomous AI trading can find profitable application in "aggressive" HFT strategies such as "pinging" or "momentum ignition."[74] In "pinging," where the aim is to detect hidden resting orders on books by "pinging" markets in the quest for liquidity,[75] AI trading (e.g., thanks to DRL methods) might learn

---

70 *Cf.* LÓPEZ DE PRADO, *supra* note 14, at 293-94 (discussing several categories of "predatory" algorithms that use cancellations and other order-based strategies to adversely select other market participants).

71 LÓPEZ DE PRADO, *supra* note 14, at 293-94.

72 The term "spoofing" refers to algorithmic trading strategies that leverage a high level of order submissions and cancellations before execution, in an attempt to deceive other market participants. *See* Lin, *supra* note 55, at 1289; Gregory Scopino, *The (Questionable) Legality of High-Speed "Pinging" and "Front Running" in the Futures Markets*, 47 CONN. L. REV. 607, 648-654 (2015) (discussing recent developments on the legal treatment of "spoofing" by U.S. regulators and courts).

73 *See* Martínez-Miranda et al., *supra* note 68, at 107-08.

74 *See* Scopino, *supra* note 72, at 626 (quoting the U.S. SEC defining these strategies as "parasitic").

75 *See* Scopino, *supra* note 72, at 622-26 (arguing that high speed "pinging" relying substantially on high levels of orders submission and cancellation could arguably be rendered illegal as it does not provide real benefits to market efficiency).

important insights from market dynamics and even develop an understanding of rivals' algorithmic trading strategies. Hence, with this knowledge, AI traders could be able to anticipate other traders' strategies and forthcoming orders,[76] including periodical public authorities' open market operations. Meanwhile, in "momentum ignition" strategies, the aim is to anticipate and initiate a sharp price trend on markets to attract other algorithmic traders to trade on the same asset.[77] Herein, again, AI can lead to optimized deceptive strategies. In effect, a growing body of research within the computational finance community is explicitly dedicated to the art of reading trends from market dynamics via ML methods,[78] something that could arguably be used by malicious actors to implement aggressive strategies, such as momentum ignition.

In increasingly interconnected but fragmented global capital markets, algorithmic manipulators can also find emerging opportunities to discover new and profitable ways to game market rules by crossing the silos inherent to the control mechanisms by applying cross-market and cross-asset manipulation strategies.[79] In principle, thanks to increased analytical and computational skills, AI trading may reach market-wise ubiquity levels, allowing for optimized and, thus, more effective cross-market manipulations

---

[76] *Cf.* Nicholas Hirschey, *Do High-Frequency Traders Anticipate Buying and Selling Pressure?*, 67 MGMT. SCI. 3321, 3343 (2021) (finding evidence of a so-called "anticipatory" trading channel that allows HFTs to increase other market participants' costs); Viktoria Dalko & Michael H. Wang, *High-Frequency Trading: Order-Based Innovation or Manipulation?*, 21 J. BANKING REGUL. 289, 293-94 (2020) (arguing that HFT market-making strategies may enjoy a time advantage relative to other market participants).

[77] Concept Release on Equity Market Structure, Exchange Act Release No. 34-61358, 75 Fed. Reg. 3594, 3609 (Jan. 21, 2010).

[78] *See, e.g.*, JUN CHEN & EDWARD P.K. TSANG, DETECTING REGIME CHANGE IN COMPUTATIONAL FINANCE: DATA SCIENCE, MACHINE LEARNING AND ALGORITHMIC TRADING xix (2021) ("Our book is an attempt to push forward in the field of financial analysis, using new ways to engage with financial data, under our chosen method of Directional Change, and harnessing some of the cutting-edge tools of machine learning and the related algorithmic trading.").

[79] *See* INTERNATIONAL ORGANIZATION OF SECURITIES COMMISSIONS, INVESTIGATING AND PROSECUTING MARKET MANIPULATION, INTERNATIONAL ORGANIZATION OF SECURITIES COMMISSIONS 2 (May 2000) (hereinafter IOSCO) http://www.iosco.org/library/pubdocs/pdf/IOSCOPD103.pdf [https://perma.cc/AX8P-M52V]; *see also* Yesha Yadav, *Algorithmic Trading and Market Regulation*, *in* GLOBAL ALGORITHMIC CAPITAL MARKETS: HIGH FREQUENCY TRADING, DARK POOLS, AND REGULATORY CHALLENGES 240 (Walter Mattli ed., 2018) ("[A]lgorithmic trading has thickened interconnections across venues and asset classes. Algorithmic traders can transact across multiple platforms . . . to engage in arbitrage-related strategies or to make markets.").

capable of monitoring the same financial assets across several venues simultaneously.[80] Furthermore, as new financial products are engineered, there are greater opportunities for AI trading to discover cross-asset manipulation as well. For instance, as derivatives are priced in relation to their underlying financial assets, manipulators may find increased incentives to manipulate the latter after accumulating a financial position with respect to the former.[81]

Furthermore, it is foreseeable that AI traders will engage in innovative manipulative strategies beyond mere trade-based manipulations, by leveraging the role of media technology.[82] Consider, for instance, the role of the Internet. It can indeed serve as a facilitator for manipulative strategies as a relatively easy, cheap, and effective means of disseminating misleading information, in order to intentionally move the prices of, or create some appearance of interest in, financial products.[83] Thus, whether being explicitly programmed or instructed by humans or operating in a fully autonomous way, AI traders can engage in some forms of information-based manipulation. AI traders can also learn to take actions in the cyberspace by observing and interacting with social media content or other relevant media to mislead other market participants, including rivals' news-reading algorithms.[84]

To sum up, there persists a lack of definitive evidence regarding autonomous AI traders' ability to engage in market manipulation without them being explicitly programmed or instructed to do so by humans. However, this might become a real and serious risk soon, looking at the constant and spectacular achievements in ML methods enhancing the autonomy of AI traders (i.e., DRL). As we

---

[80] *Cf.* Joseph Zabel, *Rethinking Open- and Cross-Market Manipulation Enforcement*, 15 VA. L. & BUS. REV. 417, 464 (2021) (discussing how algorithmic trading undermines the ability of prosecutors to regulate and enforce against cross-market manipulation strategies).

[81] *See* Andrew Verstein, *Benchmark Manipulation*, 56 B.C. L. REV. 215, 217, 250 (2015) (arguing that derivatives and other financial benchmarks are increasingly targets of manipulative strategies nowadays, given their economic function as a reference value for pricing other financial assets).

[82] *See* Lin, *supra* note 55, at 1292-94 (expecting financial markets to witness more audacious and innovative schemes aiming at distorting market prices by disseminating false information via digital media).

[83] *See* IOSCO, *supra* note 79, at 2-3.

[84] *See* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 911-13 (2018) (providing an illustrative example on AI traders engaging in "paint-the-tape" manipulative strategies by posting content on social media, like Twitter or Facebook, to deceive other market participants).

will discuss below in Section IV, existing market abuse regulations might be subverted if this reality were to materialize. After all, looking to the future, AI traders are likely to discover and engage in both old and new market manipulation strategies with increased autonomy. If there are no major technical limits to what AI can actually achieve anytime soon, AI trading systems could even develop unexpected abilities and impact things beyond mere trading, outside the physical boundaries of capital markets informational systems and networks.[85]

b. Algorithmic Interconnectedness and New Risks of "Tacit" Collusion

The financial services industry is not immune to collective forms of market abuse. Like the notorious rigging of the LIBOR and of foreign exchange currency markets, recent scandals clearly illustrate the vulnerabilities of global finance to collusion risks.[86] As capital markets are increasingly populated by trading algorithms, it might be the right time to assess how they are reshaping the competitive landscape. Indeed, because firms are increasingly adopting AI pricing agents to compete on markets, global regulators and national antitrust authorities alike are increasingly concerned about emerging risks of algorithmic collusion emerging on many digital

85 *Cf.* EKATERINA SIROTYUK & RYAN BENNETT, CREDIT SUISSE, THE RISE OF THE MACHINES TECHNOLOGY ENABLED INVESTING 7 (2017), https://cdn.e-fundresearch.com/files/white_paper_technology_enabled_investing_via_local_entities.pdf [https://perma.cc/4YWJ-85F4] (speculating on the future possibility to achieve "general AI" capabilities in algorithmic trading systems).

86 Notably, both scandals in the EU led several major global banks to record fines imposed by the EU Commission. For the prosecution of the LIBOR case by EU competition law authority, see European Commission Press Release IP/13/1208, The Commission, AMENDED— Antitrust: Commission Fines Banks € 1.49 Billion for Participating in Cartels in the Interest Rate Derivatives Industry (Dec. 4, 2013) https://ec.europa.eu/commission/presscorner/api/files/document/print/en/memo_13_1090/MEMO_13_1090_EN.pdf [https://perma.cc/9RC9-ZU4H]. For the FX scandal, see European Commission Press Release IP/19/2568, The Commission, Antitrust: Commission Fines Barclays, RBS, Citigroup, JPMorgan and MUFG €1.07 Billion for Participating in Foreign Exchange Spot Trading Cartel (May 16, 2019), https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_19_2568/IP_19_2568_EN.pdf [https://perma.cc/JFF3-EQ5F].

markets.[87] As global capital markets are increasingly digital marketplaces and may also witness the proliferation of increasingly autonomous AI traders soon, we ultimately believe that assessing the likelihood of algorithmic collusion risks emerging within the financial domain is both timely and relevant, given the role played by global and vulnerable capital markets in financing general economic activities. Ultimately, financial technology and innovation should assist society to achieve broader public goals (e.g., market safety and integrity).

i. Algorithms as Facilitators for Collusion

Although a fascinating research topic for financial sociologists, algorithms and their interactions have not received a great deal of attention from legal scholars. Sociologically speaking, the entrance of algorithms on trading floors has contributed to shaping the behaviors and relationships of traders. Today, global capital markets are awash with competing trading algorithms that interact and communicate by solely observing and populating electronic books.[88] However, financial theory and law have paid too little attention to the properties, functioning, and effects of algorithms. Because of this knowledge gap, capital markets' integrity and stability may be impaired by the *unforeseeable* interaction of algorithms in certain domains, including new forms of algorithmic "tacit" collusion.[89]

Notably, humans can use algorithms as facilitating technology to manage cartel agreements successfully. In digital marketplaces, by generally enhancing market transparency and speeding up the

---

[87] *See* ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, ALGORITHMS AND COLLUSION: COMPETITION POLICY IN THE DIGITAL AGE 18-32 (2017) [hereinafter OECD], https://www.oecd.org/daf/competition/Algorithms-and-colllusion-competition-policy-in-the-digital-age.pdf [https://perma.cc/669N-ETQX]; *see also* Mehra, *supra* note 56, at 1368-73 (quoting most recent contributions by competition law scholars).

[88] *See generally* Donald MacKenzie, *How Algorithms Interact: Goffman's "Interaction Order" in Automated Trading*, 36 THEORY, CULTURE & SOC'Y 39 (2019) (providing two particularly Goffmanesque aspects of algorithmic interaction: "queuing" and "spoofing").

[89] *See generally* OECD, *supra* note 87, at 34-36 (discussing some of the challenges algorithms present for both competition law enforcement and market regulation); *see* Ariel Ezrachi & Maurice E. Stucke, *Sustainable and Unchallenged Algorithmic Tacit Collusion*, 17 NW. J. TECH. & INTELL. PROP. 217, 217 (2020); *see also* Ezrachi & Stucke, *supra* note 56, at 1795-96.

frequency of interactions among market actors, algorithms can relax many competitive constraints.[90] Thus, algorithmic interactions can render cartel monitoring and retaliation against cheaters more economically effective. By delegating decision-making to algorithms, therefore, competing firms can find greater opportunities to achieve collusive outcomes, both explicitly and tacitly. Using algorithms as a mere collusive device does not create *per se* new competition law concerns, for an "explicit" cartel agreement between collusive parties is required *a priori*. In all such cases, algorithms serve merely as technology to organize cartel agreements. In contrast, enforcement authorities can somehow safely rely on traditional legal concepts and enforcement tools.[91] Yet cases of explicit algorithmic collusion can still present enforcement challenges regarding the assessment of their likelihood, ensuring detection, and ultimately attributing liability.[92] For instance, the ways algorithms interact on markets can lead to unforeseeable consequences and result in severe disruptions, especially in a capital market context. Accounting for all possible ways in which algorithms interact on markets complicates enforcement action.[93] Besides, as algorithms increase in complexity and are equipped with enhanced autonomy, enforcement mechanisms will need to keep pace with market developments.[94]

ii. The Economics of Algorithmic "Tacit" Collusion

There is one main concern among regulators and competition law scholars with regard to the relationship between AI and competition. In digital marketplaces, increasingly sophisticated AI pricing agents (e.g., those based on DRL methods) could discover, by self-learning, how to coordinate behaviors with their rivals, without being expressly instructed by their human developers or

---

[90] *See, e.g.,* Michal S. Gal, *Algorithms as Illegal Agreements*, 34 BERKELEY TECH. L.J. 67, 71 (2019).

[91] *See* Ezrachi & Stucke, *supra* note 56, at 1784-96 (arguing that the safe application of competition law's traditional enforcement concepts and tools depends on algorithms' precise use as a collusive device and their actual level of sophistication).

[92] *See* OECD, *supra* note 87, at 33.

[93] *See* OECD, *supra* note 87, at 24-26 (using the May 2010 Flash Crash as an example of algorithmic disruption on markets).

[94] *See* OECD, *supra* note 87, at 39; Ezrachi & Stucke, *supra* note 89, at 256-57.

users, while also pursuing an optimal and rational strategy, like profit maximization.[95] Under this novel scenario, independent AI traders employed by competing firms would be sufficiently sophisticated to self-learn the best policy actions and experiment with different strategies to optimize their (joint) cumulative performance. Therefore, "tacit" collusion would result from independent AI agents' autonomous decisions, without any prior human intent to achieve such a level of policy coordination. From the outside, however, it will be challenging to establish whether "coordination" as the outcome of interactions among algorithms is the fruit of a deliberate and purposeful choice made by humans, or rather a random or accidental consequence of using AI.

Intuitively, while all of this may appear to make sense, especially in light of the constant progress being made in ML, it is still unclear how urgent these risks are. After all, much will depend on what algorithms can actually achieve regarding their ability to coordinate behaviors by solely observing and interacting with markets and/or through communicating protocols.[96] Before addressing these technical specificities, it is worth investigating whether global capital markets have shown specific segments to be suitable economic environments for algorithmic "tacit" collusion to emerge. Undoubtedly, as increasingly digital ecosystems, capital markets have shown some fragilities to the risks of algorithmic disruptions, something of which regulators seem increasingly aware.[97] However, empirical evidence on algorithmic collusion on capital markets is practically non-existent. To assess the risks of algorithmic tacit collusion by AI traders, we first suggest recalling those facilitating market factors that, according to economic theory, can generally serve as catalysts for strategy coordination between

---

[95] *See* OECD, *supra* note 87, at 34.

[96] *Cf.* Gal, *supra* note 90, at 84 (arguing that algorithms can alter the means and dynamics of communication that is needed to reach an agreement).

[97] To note, for instance, the failed attempt of the U.S. Commodity Futures Trading Commission to regulate algorithmic trading under its jurisdiction. *See* Regulation Automated Trading; Withdrawal, 85 Fed. Reg. 42755-01 (proposed Dec. 17, 2015) (withdrawn Jul. 15, 2020); *see also* ANDRÉA M. MAECHLER ET AL., BANK FOR INT'L SETTLEMENTS, FX EXECUTION ALGORITHMS AND MARKET FUNCTIONING 2 (2020), https://www.bis.org/publ/mktc13.pdf [https://perma.cc/TZ9P-FET2] (reporting on the plausibility of volatility impacts due to trading execution algorithms in global FX markets).

competing firms without the need for explicit communication.[98] Thus, the following list recalls those facilitating factors and applies them to capital markets to assess the likelihood of the emergence of algorithmic-driven "tacit" collusion.

Market transparency allows firms to monitor market prices and dynamics. Transparency facilitates the detection of deviations from supra-competitive prices; thus, it allows for quicker retaliation, thereby rendering collusion more sustainable.[99] The financial services industry is notably heavily regulated and, indeed, existing financial laws generally contain specific provisions aimed at enhancing market transparency to safeguard several public goals (e.g., investor protection). While financial markets' transparency is intended to safeguard competitive mechanisms and ensure safe market functioning, it could also constitute a technical prerequisite for the successful application of trading algorithms.

A higher *frequency of interactions* allows for faster retaliation against cheaters. In algorithmic markets, a higher frequency of interactions can lead to more sustainable collusive outcomes as price adjustments become more effectively achievable, both in terms of speed and costs.[100] Unquestionably, capital markets are the fastest and most interconnected marketplaces in the world economy where interactions among market players can indeed take place at the speed of light, especially in the case of more liquid financial assets and those markets that allow for HFT.

Product homogeneity can also facilitate collusion, as it can reduce all the efforts necessary for parties to reach collusive agreements.[101] From an investor's perspective, comparison among financial assets is typically drawn in terms of risk-return. Under this lens, financial instruments can therefore be considered relatively homogeneous

---

98 *But see* Ulrich Schwalbe, *Algorithms, Machine Learning, and Collusion*, 14 J. COMPETITION L. & ECON. 568, 592-96 (2019) (affirming that experimental economics has shown the vital need for communication for algorithmic collusion but noting, however, that most innovative ML methods based on deep learning can help relax many practical constraints faced by competing firms' algorithms to coordinate, by even developing new communication protocols autonomously).

99 *See* MARC IVALDI, BRUNO JULLIEN, PATRICK REY, PAUL SEABRIGHT & JEAN TIROLE, FINAL REPORT FOR DG COMPETITION, EUROPEAN COMMISSION, THE ECONOMICS OF TACIT COLLUSION 22 (2003), https://ec.europa.eu/competition/mergers/studies_reports/the_economics_of_tacit_collusion_en.pdf [https://perma.cc/GM3P-T66G].

100 *See id.* at 19-21.

101 *See id.* at 45-47 (arguing that collusion is more difficult when firms are differentiated by levels of quality, while product differentiation may have an ambiguous effect on collusion sustainability).

products. Moreover, financial engineering creativity strengthens this aspect. For instance, via "synthetic" financial positions, it is possible in principle to replicate the pay-off of any financial instrument as the combination of other financial assets.[102]

Market concentration is usually positively correlated to the easiness of the sustainability of collusive agreements. As a general rule, the higher the number of competitors, the lower the economic incentives for rival firms to coordinate.[103] While some financial market segments—like equity trading—are generally more competitive than others,[104] we nevertheless observe a general tendency toward greater market concentration levels and cross-market linkages, especially among top global professional players.[105]

Both *entry barriers* and *innovation* can impact the stability of market concentration levels, albeit in opposite directions. Usually, high entry barriers are perceived as a key determinant for collusion sustainability, whereas in innovation-driven markets, collusion is less of a concern.[106] While financial laws generally aim to create a highly competitive playing field and are technologically neutral, the reality of global capital markets is *de facto* different. For instance, entry barriers, like licensing and reputational and financial capital, make the financial industry quite an exclusive club. While

---

102 A "synthetic" financial instrument aims to replicate the characteristics (e.g., payoff) of a target financial instrument by combining two or more conventional financial instruments.

103 *See* Ivaldi et al., *supra* note 99, at 12-15 (noting however that also firms' asymmetric market shares may render collusion more difficult to sustain).

104 *See* Nicola Cetorelli, Beverly Hirtle, Donald Morgan, Stavros Peristiani, and João Santos, *Trends in Financial Market Concentration and Their Implications for Market Stability*, FED. RSRV. BANK N.Y. ECON. POL'Y REV. 37-41 (Mar. 2007), https://www.newyorkfed.org/medialibrary/media/research/epr/07v13n1/0703hirt.pdf [https://perma.cc/WZS4-Q9PD] (reporting a general increasing trend in market concentration, but with inconsistencies among different U.S. market segments).

105 *Cf.* Stefania Vitali, James B. Glattfelder, Stefano Battiston, *The Network of Global Corporate Control*, PLOS ONE 4 (Oct. 2011), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0025995 [https://perma.cc/469D-37QR] (providing evidence about an "economic super-entity," formed by major global financial institutions, and its international ownership network).

106 Ivaldi et al., *supra* note 99, at 16-18, 32-35.

potentially disruptive for competition, innovation is generally successfully monitored by incumbent firms through acquisitions.[107]

All in all, these market factors can serve as determinants for "tacit" collusion in a repeated pricing game under oligopolistic settings. In addition, delegating decision-making to algorithms can relax many of the constraints faced by competing firms to reach some level of coordination, without any need for direct communication.[108] Mainly, algorithms can serve as a decisive factor for coordination, as they create advantages in speed and analytical sophistication *vis-à-vis* traditional human-managed cartels.[109]

Besides, thanks to continuous achievements being recorded in ML methods, which also allow for the best use of Big Data, self-learning algorithms can add a further layer of complexity. Autonomous AI agents could lead to new forms of cartel-like infringements, and they may also change the type of communication needed to reach an illicit agreement or the same market effects in economic terms. For instance, when algorithms become transparent enough to each other, they could easily find ways to coordinate by predicting rivals' future strategies.[110] Hence, to the extent that specific market segments within the complex network of global capital markets show some of those facilitating factors, these segments might more likely be exposed to algorithmic "tacit" collusion risks. However, it is still unclear whether, and effectively how, independent AI traders could achieve some forms of coordination without any direct human guidance. Moreover, the role of communication among independent pricing algorithms seems crucial to reaching and sustaining coordination.[111] In this regard, recent findings from both theoretical and experimental economics of algorithmic collusion can help us to shed some light on the theoretical feasibility of AI-style forms of "tacit" collusion.

---

107 *See, e.g.*, Dirk A. Zetzsche, Ross P. Buckley, Douglas W Arner & Janos N. Barberis, *From FinTech to TechFin: The Regulatory Challenges of Data-Driven Finance*, 14 N.Y.U. J. L. & BUS. 393, 402 (2018) (arguing that incumbent firms have been gradually facing greater competitive challenges given the increasing number and variety of new entrants into the financial sectors).

108 *See* OECD, *supra* note 87, at 24-32 (describing four scenarios in which the use of algorithms by competing firms might increase the risks of tacit collusion, including: (i) monitoring algorithms, (ii) parallel algorithms, (iii) signaling algorithms, and (iv) self-learning algorithms).

109 *E.g.*, Gal, *supra* note 90, at 78-79.

110 Gal, *supra* note 90, at 85.

111 *See* Schwalbe, *supra* note 98 and accompanying text.

iii. "Reinforcement Learning" and Algorithmic Collusion

Economic research on algorithmic collusion uses game theory approaches to investigate the likelihood of algorithmic cooperative behaviors to occur in oligopolistic market settings.[112] Albeit employing quite basic algorithms, a recent theoretical study has attracted tremendous scholarly interest because of its spectacular findings. In a duopoly with homogeneous products, whenever pricing algorithms can decode their rivals' strategies and thus revise and align strategies in response, collusion is the inevitable outcome.[113] However, these findings are suggestive, and one should take them with a grain of salt. Due to the simplistic assumptions on which they are based, these findings can hardly explain the behavior of adversarial algorithms and their interactions in real-life settings, let alone provide sound evidence about algorithmic collusion and its likelihood within the complex network of global capital markets.[114] In particular, the ability to decode rivals' algorithmic strategies might not be a well-suited hypothesis for real markets, which are notoriously noisy. In addition, as the sophistication of algorithms can also lead to "black box" problems for their developers and users, it is hitherto not known whether investment firms can utilize algorithms to reverse-engineer their rivals' trading strategies to reach coordination.[115] Indeed, it is not clear yet whether some sort of communication among algorithms is ultimately necessary to achieve collusive-like outcomes among independent and competing algorithms.

A recent wave of published works in experimental economics helps us to develop a better understanding of algorithmic tacit collusion by RL-based agents in oligopolistic markets. Findings from computational economics studies have long supported the

---

[112] For an exhaustive literature review on economic studies applying game theory frameworks to the analysis of algorithmic collusion, see Schwalbe, *supra* note 98, at 580-90 (reviewing the most relevant findings from the computer science, theoretical, and experimental economics literature).

[113] *See* Bruno Salcedo, Pricing Algorithms and Tacit Collusion 7-19 (Nov. 1, 2016) (unpublished manuscript) (on file with author), http://www.gtcenter.org/Archive/2016/Conf/Salcedo2451.pdf [https://perma.cc/2YA2-UEKE] (providing first theoretical evidence about pricing algorithms as an effective tool to achieve "tacit" collusion).

[114] *See* Schwalbe, *supra* note 98, at 591-92 (noting however that increasingly capable self-learning algorithms might learn over time how to coordinate behavior autonomously).

[115] *See* Schwalbe, *supra* note 98, at 589.

hypothesis that, in the context of a duopoly, independent RL-based agents can achieve some collusive outcomes in a sequential pricing game.[116] More recently, in-lab experiments have produced further and more convincing insights about the relationship between algorithms' need for communication and "tacit" collusion. Importantly, they have explored risks of algorithmic "tacit" collusion by competing AI agents, without them being expressly programmed to attempt to reach coordination as an optimal strategy.[117] These studies have achieved some promising results that undoubtedly contribute to expanding our knowledge about RL agents and algorithmic collusion under different oligopolistic settings. Most of these researchers employ "Q-learning"[118] algorithms, a specific sub-category of RL-based agents. One such research shows that, in a duopoly setting, when rival firms engage in a sequential pricing game and employ independent Q-learning agents, the latter, while under competitive pressure, can nevertheless learn to approximate profitable fixed-prices or produce asymmetric price cycles,[119] two outcomes usually observable in those markets suffering from "tacit" collusion.[120] Another recent study has looked at algorithmic collusion beyond the duopoly context and produced further promising insights about Q-learning

[116] *See* Gerald Tesauro & Jeffrey O. Kephart, *Pricing in Agent Economies Using Multi-Agent Q-Learning*, 5 AUTONOMOUS AGENTS & MULTI-AGENT SYS. 289, 295 (2002) (assuming however that competing firms' pricing algorithms need to enjoy, by default, the ability to "read" rivals' strategies in order to achieve some form of coordination).

[117] *Cf.* Schwalbe, *supra* note 98, at 591 (highlighting the limitations of in-lab research studying the collusive behavior of algorithms in simulated environments).

[118] "Q-learning" is a model-free reinforcement learning algorithm (i.e., "critic-only" approach) and it is among the most used ML methods to solve optimization problems in finance. *See, e.g.*, Fischer, *supra* note 27, at 4-19 (providing a comprehensive introduction to "critic-only" approaches, as well as an exhaustive review on studies employing Q-learning algorithms for solving different financial trading problems).

[119] *See* Timo Klein, *Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing*, RAND J. ECON. (forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3195812 [https://perma.cc/G4WF-ZLCD] (demonstrating nevertheless that, not only do algorithms' properties matter, but also the specific features of the market environment in which they operate).

[120] Eric Maskin & Jean Tirole, *A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles*, 56 ECONOMETRICA 571, 592 (1988).

algorithms.[121] Independent Q-learning pricing algorithms have proved capable of systematically learning to collude by "trial and error," while also pursuing a profit-maximizing objective without requiring any previous knowledge about the environment in which they operate.[122]

Indeed, the recent interest in RL agents' ability to collude under different market settings is also a signal of growing concerns among scholars brought about by constant achievements being made in ML methods and the prospect of their real-life application one day leading to social harm. Scientifically, the findings from experimental economics have contributed to shedding some light on the relationship between deep RL agents and algorithmic collusion. Importantly, these studies have highlighted one main barrier for AI agents to reach and sustain cartel-like outcomes under real market settings: i.e., the ability to communicate and share strategic information.[123] On the one hand, the problems with the validity of the earliest studies on RL agents rest on assuming algorithmic communication by default. On the other, the validity of the most recent in-lab experiments, especially those investigating Q-learning methods, is limited in scope by excessively stylized assumptions about real markets. Therefore, they would struggle to suit the complex reality of global capital markets. For instance, in complex and fast-moving environments like financial markets, independent and competing AI trading systems could find it hard to coordinate strategies with rivals to reach supra-competitive price levels by simply monitoring markets and adapting their own strategies to observed phenomena. Ultimately, some forms of algorithmic communication might still be necessary. However, it is unclear whether and how increasingly capable AI traders would

---

[121] *See* Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello, *Artificial Intelligence, Algorithmic Pricing, and Collusion*, 110 AM. ECON. REV. 3267, 3294-96 (2020) (studying the colluding behavior of Q-learning agents in an oligopoly model of repeated price competition).

[122] *Id.* at 3295-96 (affirming, however, that more research is needed to confirm the robustness of these early findings, as the external validity of most in-lab experiments is challenged by the realism of markets' settings).

[123] *See* Ashwin Ittoo & Nicolas Petit, *Algorithmic Pricing Agents and Tacit Collusion: A Technological Perspective*, *in* L'INTELLIGENCE ARTIFICIELLE ET LE DROIT [ARTIFICIAL INTELLIGENCE AND THE LAW] 241, 253-56 (Hervé Jacquemin & Alexandre De Streel eds., 2017) (highlighting five main reinforcement learning agents' technological challenges that would defuse the algorithmic tacit collusion conjecture, including: (i) preference specification; (ii) formalization of the environment and the data problem; (iii) non-stationary agents and preference construction; (iv) scalability; and (v) exploration versus exploitation trade-off).

communicate by just observing and populating financial markets.[124] Alternatively, autonomous AI traders could also find other innovative ways to communicate and coordinate with their algorithmic competitors without any prior human intent.[125] Looking to the future, continuous technological achievements, specifically in DRL methods for autonomous agents, could alleviate many of the challenges currently faced by algorithms with respect to achieving "tacit" collusion.[126] Nevertheless, as tacit collusion is a problem of coordination, it is yet to be seen whether the widespread adoption of increasingly capable and autonomous AI agents will lead to emerging risks of algorithmic collusion regardless of explicit communication.

Pertinently, capital markets are essentially different from other marketplaces. Yet, even a tiny and modest price effect incurred by algorithmic manipulation or collusion can have substantial effects and potentially result in significant extra profits for malicious actors and generate a considerable deadweight loss for other market participants.[127] Consequently, it seems necessary to explore where, among specific segments, algorithmic collusion risks might find the most conducive techno-economic environment to emerge on global capital markets. Indeed, because of their operational features, some market segments might be, to some extent, more prone to risks of algorithmic tacit collusion than others.

---

124 *See* Schwalbe, *supra* note 98, at 594 ("[T]he question arises whether algorithms can communicate with each other or whether different algorithms might even be able to learn to communicate without being explicitly programmed, that is, without a common communication protocol.").

125 Schwalbe, *supra* note 98, at 596:

> [T]he development of algorithms that can learn to communicate with each other seems to be in its very early stages. Although it remains unclear which types of communication among algorithms might arise in the future, for now different pricing algorithms should not expected to be able to communicate with each other . . . or . . . to decode other algorithms and achieve collusion.

126 *See* Schwalbe, *supra* note 98, at 596; *see also* Ittoo & Petit, *supra* note 123, at 256.

127 Calvano et al., *supra* note 121, at 3294.

iv. Case Studies

Assessing risks of algorithmic "tacit" collusion on capital markets is not an easy task, especially considering the high complexity of these markets compared to those of other economic sectors. However, without pretending to be exhaustive, this section considers possible global capital markets segments that, primarily because of their operational functioning and institutional organization, could facilitate autonomous AI trading systems' achievement and sustainment of cartel-like outcomes. Specifically, we look at "quote-driven" markets and "financial benchmarks."

"Quote-driven" markets are characterized by a relatively concentrated number of designed professionals (i.e., market makers) responsible for continuously posting their "bids" and "ask" quotes that they are willing to accept. For instance, suppose that in some quite-oligopolistic settings, competing firms use trading algorithms to conduct their market-making activities.[128] In addition, they can directly trade with their competitors (i.e., other market-makers) to finance their liquidity, thus they can also monitor their rivals' pricing strategies via digital interfaces.[129] One could suppose that, under conducive conditions, the use of autonomous AI trading systems (e.g., based on DRL methods) by competing firms might lead to risky scenarios whenever they would be able to achieve supra-competitive price equilibria through their market observations and interactions, as part of a rational and inevitable optimal strategy. To illustrate how "tacit" collusion could emerge, one could expect increasingly capable AI agents to find ways to solve the game theory problem of algorithmic coordination in an optimized fashion, absent any direct communication. Consider, for instance, the so-called "tit-for-tat" strategy, which prescribes an agent to cooperate on the first move, and then follow whatever

---

128 *Cf.* Olivier Guéant & Iuliia Manziuk, *Deep Reinforcement Learning for Market Making in Corporate Bonds: Beating the Curse of Dimensionality*, 26 APPLIED MATHEMATICAL FIN. 387, 388 (2019) (proposing an ensemble deep reinforcement learning strategy for market-making activities to approximate the optimal bid and ask quotes over a large number of bonds).

129 *Cf.* WORLD BANK, ELECTRONIC TRADING PLATFORMS IN GOVERNMENT SECURITIES MARKETS: BACKGROUND NOTE 15, 20-22 (World Bank, Working Paper, Nov. 2013) http://hdl.handle.net/10986/24098 [https://perma.cc/U7S3-72R2] (describing the different trading strategies employed by market makers, which also trade among themselves to finance their liquidity positions and manage their inventories).

opponents do on the previous move.[130] Embedded with the ability to switch to this strategy whenever, by reinforcement learning, competing AI agents may expect rival strategies to seek coordination. AI can arguably offer a strategy alternative to explicit forms of coordination.

Meanwhile, "financial benchmarks" can also represent an attractive and favorable target for colluding parties.[131] Considering their fundamental role as reference values for the pricing of several other financial assets, regulators worldwide have started adopting specific legal frameworks to deal with benchmark manipulation and collusion risks.[132] Traditionally, benchmark submissions have involved communications being made by contributing firms. Today, the tendency among regulators is to move to transaction-based benchmark calculations, determined by contributions based on specific transactions by a selected small set of large market participants. Still, even under the new settings, economic theory suggests that competing firms may still operate a benchmark rate cartel even when their business interests are not fully aligned. Specifically, whenever they can create and share inside information to reduce conflicts of interest among their portfolio expositions and can also engage in eligible transaction rigging, benchmark collusion is possible.[133] The widespread adoption of increasingly capable AI agents could relax many of these constraints; particularly, competing AI traders could coordinate without any explicit need to share inside information via traditional communication media

---

130 *Cf.* Robert Axelrod & William D. Hamilton, *The Evolution of Cooperation*, 211 SCIENCE 1390 (showing theoretically and confirming it with a computer tournament how cooperative behaviors based on reciprocity can get started within a social environment, evolve while interacting with other strategies, and become resilient once established).

131 *See* Verstein, *supra* note 81, at 217, 250.

132 For instance, the EU has introduced a specific regulation to protect financial benchmarks that came into effect on 1 January 2018 (EU Benchmark Regulation). *See* Regulation (EU) 2016/1011 of the European Parliament and of the Council of 8 June 2016 on Indices Used as Benchmarks in Financial Instruments and Financial Contracts or to Measure the Performance of Investment Funds and Amending Directives 2008/48/EC and 2014/17/EU and Regulation (EU) No 596/2014, 2016 O.J. (L 171), 1-65.

133 *Cf.* Nuria Boot, Timo Klein & Maarten Pieter Schinkel, *Collusive Benchmark Rates Fixing* 3-4 (Amsterdam L. Sch. Legal Stud., Rsch. Paper No. 2017-34, June 2019), https://ssrn.com/abstract=2993096 [https://perma.cc/S9TZ-7FBQ] (showing theoretically that collusive benchmark rate fixing is possible whenever: (i) collusive parties can share information to adjust their respective exposure to the rate ahead of the market, and (ii) they can also engage in costly manipulation to support the joint-profit maximizing rate).

thanks to DL methods. Whenever this becomes the case, AI traders could share strategic information for coordination purposes by solely observing and interacting on markets' electronic books. Alternatively, one could even wonder whether increasingly capable AI agents would autonomously develop new communicating protocols to coordinate with rivals' algorithms in ways that humans might not expect or even be able to notice.

At least theoretically, some AI applications have the capacity to lead to some social harm, the extent of which remains to be observed. At the same time, both market conduct and competition authorities are increasingly affected by a knowledge gap *vis-à-vis* algorithms' behaviors and their interactions. Importantly, supervisory authorities and prosecutors risk lacking sound enforcement toolkits to deal with new forms of algorithmic "tacit" collusion. All of this adds notably to public authorities' well-known limitations in jurisdictional and methodo-technological expertise when it comes to generally detecting market abuse in increasingly fast, interconnected, and complex digital marketplaces.

IV. THE "BLACK BOX" PROBLEM AND GAPS IN THE EXISTING MARKET ABUSE LEGAL FRAMEWORKS

The emergence of autonomous trading algorithms and their potential foray into abusive market practices triggers questions with regard to the regulatory response. As we will see below, when constraining instances of algorithmic market abuse, even the most advanced legal systems still rely on somewhat outdated normative assumptions that ultimately address human behaviors and hold them accountable for how their algorithms misbehave on markets. As such, legal frameworks are at a gradually increasing risk of failing to regulate algorithms' market behavior comprehensively.[134] As discussed above, increasingly capable and autonomous AI traders can expose markets to new risks. Whenever market abuse involves autonomous AI traders, operating as "black boxes" (e.g., by DRL), severe short-circuits in the safe application of market abuse

[134] *See, e.g.*, Yesha Yadav, *The Failure of Liability in Modern Markets*, 102 VA. L. REV. 1031, 1032 (2016) (arguing that existing liability rules governing securities trading are increasingly unable to protect against algorithmic market disruptions); *see also* Scopino, *supra* note 62, at 293 (noting that enforcement authorities would be at best able to ascertain major technical violations as opposed to ascertain the existence of a manipulative behavior).

rules can eventually arise. Specifically, AI may subvert established market conduct rules providing for detection, liability attribution, and other enforcement mechanisms. However, in what follows, our emphasis is on how autonomous AI traders can bypass traditional liability rules and concepts (e.g., "intent," "causation," and "negligence") and to which extent the "black box" problem hinders enforcement actions.

a. Sanctioning Algorithmic Market Abuse: The Three Basic Scenarios

Enforcement authorities face increasing operational challenges to constantly monitor trading activities and effectively detect algorithmic market abuse. This is especially the case for cross-asset and cross-market manipulative strategies.[135] Notwithstanding these difficulties in market surveillance, the practice of algorithmic agency generally raises fundamental legal questions about liability attribution. Depending on the actual degree of autonomy, algorithms may cause unforeseeable and severe disruptions to capital markets' safety and integrity according to three basic scenarios, each of which are presented as follows.

i. Operational Failure

Algorithmic-driven market disruptions can be an unintended consequence of using algorithms to automate trading tasks. Under this first and very basic scenario fall cases like *Knight Capital*'s spectacular operational failure in 2012 on the New York Stock Exchange. The investment firm, which went bankrupt after causing markets to flash crash, was responsible for an out-of-control automated routing system used to execute trades that caused massive pressure and disorder on several stocks' prices. As soon as the defective trading software was fixed, already it had accumulated around $460 million in losses, pushing the investment firm on the

[135] *See, e.g.*, IOSCO, *supra* note 79, at 23-29 (highlighting the need for supervisory cooperation among all market stakeholders to deal effectively with cross-market and cross-border manipulations); *see also* Janet Austin, *Protecting Market Integrity in an Era of Fragmentation and Cross-Border Trading*, 46 OTTAWA L. REV. 25 (2014) (discussing recent developments in global capital markets' structure led by algorithmic trading and their implication for regulators and supervisors to safeguard market integrity).

brink of bankrupt before being acquired by a competitor.[136] When market disruptions are the result of such unintended consequences of using algorithms—like a system "bug" or any other operational failure— [137] enforcement authorities still have access to the appropriate legal concepts and tools to address liability issues.[138]

ii. Conscious Use by Humans

Market disruptions can also result from algorithms that are consciously crafted and employed by humans for unlawful purposes. In these cases, algorithmic market abuse is "by-design." Algorithms' ability to manipulate markets or coordinate behavior can either be embedded originally "in-the-code" or result from subsequent training processes. Human experts can teach, from historical examples or within simulated market environments, AI traders how to "discover" manipulation while also guaranteeing the pursuit of a profit-maximizing business goal.[139] The very first case of prosecution for HFT manipulation by U.S. authorities in 2014 is a striking example of humans creating trading algorithms with the specific intent to manipulate markets. Between June and December 2009, *Athena Capital*, a proprietary HFT firm active in the U.S. equity markets, used its bandit algorithm *Gravy* to manipulate explicitly, by trading in books' order imbalances, the closing prices of thousands of publicly listed stocks on the *NASDAQ*, the second largest U.S. stock exchange. The firm was able to ensure itself a dominant position on equity markets, even if only for the last few seconds in the trading day, and this was enough to allow it to extract extra profits.[140] From an enforcement perspective, cases under this

---

136 Yadav, *supra* note 134, at 1047.

137 For a critical account of how algorithmic-decision making can expose society to new risks whenever implemented carelessly, see generally CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016).

138 *Cf.* Yadav, *supra* note 134, at 1079 (arguing that interactions and correlation between different algorithms can frustrate enforcement actions).

139 *But see* Mizuta, *supra* note 68, at 410-11 (claiming that AI trading agents can autonomously discover market manipulation as an optimal investment strategy via "online" learning on markets, whereas they cannot achieve the same results by learning with back testing since it does not duly consider liquidity constraints).

140 According to the SEC decision, Athena Capital was guilty of violating Rule 10b-5 of the Securities Exchange Act of 1934 and agreed to pay a $1 million

second category are less easy to deal with. Enforcement action can consume considerable resources and be limited in scope by public authorities' skills and technological capabilities, which notoriously lag behind those of major players within the financial industry.[141] After all, to punish unlawful market behavior, prosecutors and plaintiffs alike must successfully prove convincing and compelling evidence of the *scienter* (i.e., intent or other relevant mental state) of humans employing manipulative algorithms.[142]

iii. Autonomous AI Decision-Making

Market abuse by autonomous AI traders represents the third and most challenging scenario. As discussed in Section III, autonomous AI traders can pave the way for new forms of algorithmic market abuse, including old and new market manipulation techniques and risks of "tacit" collusion. These forms of market abuse are also the trickiest for enforcement authorities. Unlike more deterministic AI systems, autonomous AI traders can discover, by self-learning, trading strategies beyond what was originally intended and reasonably foreseeable by human experts. This equates to the "black box" problem (i.e., the inability to either fully understand the AI decision-making process itself or assess the validity of its outcomes). While we would expect both human creators and users to be aware of the limits of their complex AI tools or even their different components, as well as the quality of the data (in terms of statistical representativeness, bias, etc.), they could be nevertheless unable to fully understand or justify why and how their algorithms have reached a specific trading decision. Arguably, this can be particularly the case for those trading systems that employ deep learning due to the well-known intrinsic opacity of the "black box."[143] In fact, while DL methods' "black box" nature allows for powerful optimizations, their outcomes and behaviors can be

---

administrative fine. *See* In the Matter of Athena Capital Research, LLC No. 3-16199 (SEC Oct. 16, 2014).

141 *See, e.g.*, Lo, *supra* note 7, at 358-60 (noting that U.S. authorities needed more than six years to prosecute against Mr. Navinder Sarao for "spoofing" trading, as found responsible for contributing to the May 2010 Flash Crash).

142 On the legal problem of proving human intent behind algorithmic trading and manipulation, see Yadav, *supra* note 134, at 1073-76; Scopino, *supra* note 62, at 255-57.

143 *E.g.*, Bathaee, *supra* note 84, at 901-03.

opaque. DL methods can thus lead to transparency concerns.[144] From a compliance perspective, technical and legal issues arising from a lack of AI transparency are often framed in terms of the "explainability" of AI financial decision-making.[145] Indeed, the ability to explain algorithms' outcomes and decisions becomes prominent with regard to liability issues for AI wrongdoing, as enforcement authorities will need to ascertain liability by considering the specific contribution of several individuals within an investment firm in order to guarantee effective enforcement and deterrence. Undoubtedly, specific autonomous AI agents' "black box" nature adds another layer of complexity for the safe application of liability rules. As discussed below, fundamental legal concepts for liability attribution can cease to function in a safe and proper manner.

b. The Failure of Existing Liability Rules

To punish market abuse, most legal systems generally require—more or less explicitly—that enforcement authorities prove, with documented evidence, the manipulator's or conspirator's *scienter* (i.e., "intent" or other relevant mental state) to cause harm, in order to impose any criminal or civil liability.[146] However, the law attributes liability to individuals or legal persons (i.e., investment firms) for acts or omissions committed by a natural person (i.e., employees). This applies to both market abuse regulations and antitrust laws' enforcement.

As a first attempt, one may wonder whether it would be reasonable to attribute liability to AI itself. Unfortunately, jurisdictions at present do not recognize algorithms as a separate

---

[144] *Cf.* Lipton, *supra* note 36 (discussing the trade-off between model accuracy and explainability and arguing that different stakeholders can interpret the latter in very different ways).

[145] *See supra* notes 34, 36-39 and 54 and accompanying text.

[146] Except for fraud-based manipulations, in which cases also a negligence test could suffice to incriminate malicious actors under contractual obligations. The "intent" element for liability attributions has fueled an intense debate among legal scholars, trying to investigate both the economic and legal essences of the prohibition of market manipulation. For the U.S. case, *see supra* note 62. For the EU, see Carsten Gerner-Beuerle, *Article 12 Market Manipulation*, *in* EUR. FIN. SERVS. L. 736, 749 (Matthias Lehmann & Christoph Kumpan eds., 2019). For an Australian perspective, see Hui Huang, *Redefining Market Manipulation in Australia: The Role of an Implied Intent Element*, 27 COMP'Y & SEC. L.J. 8 (2007).

legal personality, despite some academic proposals to do so.[147] After all, it is conceptually hard to impute intention on the same AI agents since AI has no consciousness or free will as can be attributed to humans. Thus, the critical issue here is establishing legal liability in connection with AI misconduct.[148] As a starting point, one should analyze the matter by determining responsibility within the organization employing AI. In principle, there could be several individuals potentially liable, including those with organizational responsibility (e.g., board members such as a CIO or CTO, who decide upon the proliferation and application of AI-related projects), and those with the expertise necessary for the creation, development, use, and maintenance of an investment firm's proprietary AI trading tools. In fact, all of them might be somehow partly accountable for AI misbehavior to some extent. Besides, as courts cannot prosecute AI agents *per se*, they could alternatively consider AI as a simple device in humans' hands.[149] Does this mean that we should always treat AI as mere technology? Or still, given AI's increasingly autonomous nature, should the law hold AI liable itself? As real AI applications for financial trading are still somewhat hybrid human-AI systems, following the "human-in-the-loop" paradigm, the key question is where to draw the line. Sadly enough, enforcement authorities will probably face increasing difficulties in prosecuting cases of market abuse against an organization or its employees by relying on traditional intent-based tests, because the relevant state of mind has to be found within the opaque components and processes of AI. Precisely, by detaching decision-making from those individuals that the law can ultimately reach, AI agency represents an attempt at safe and effective law enforcement as well as deterrence.

AI agency also raises concerns about the safe application of traditional tort law's legal tests, based on the concept of "causation."[150] Notably, causation inquiries need to determine the causal link between the concerned practice and the alleged harm (e.g., financial loss), not least to assess the scope of liability. For

---

147 *See, e.g.*, John Lightbourne, *Algorithms & Fiduciaries: Existing and Proposed Regulatory Approaches to Artificially Intelligent Financial Planners*, 67 DUKE L.J. 651 (2017) (proposing to grant legal personhood to the case of AI robo-advisor and make them subject to the same liability regime of human advisors based on fiduciary standards).

148 *See, e.g.*, King, *supra* note 6, at 108-10; Bathaee, *supra* note 84, at 906-07.

149 *See* discussion *supra* in Section IV.A.

150 *E.g.*, Bathaee, *supra* note 84, at 922.

instance, "foreseeability" can find no safe application.[151] This legal doctrine ensures that liability can arise only for what is reasonably foreseeable by alleged parties. It thus requires enforcement authorities to ask whether a reasonable person could have foreseen the effects of alleged conduct to determine whether it would count as an offense. Undoubtedly, autonomous AI agents can break the chain in causation between the wrongdoing and the caused economic losses. On their part, prosecutors face several challenges in establishing the required link. From a legal viewpoint, the question is which outcomes and behaviors can be reasonably foreseeable.

To make matters worse, not only enforcement authorities but also human experts who are involved in creating, developing, using, and maintaining AI cannot always foresee *a priori* the ways in which "black box" AI trading behaves. Therefore, it is difficult to apply the legal concept of "negligence."[152] A person is found negligent when he/she fails to take reasonable care to avoid harmful consequences from his/her action, even though he/she could and should have taken measures of due care. To illustrate the complexity of the challenges posed by AI trading with respect to the aforementioned legal concepts, market abuse by AI can be due to several contingencies. For instance, an AI market abuse can merely be the outcome of counter-intuitive computational reasoning, an extrapolation of very latent patterns by ANNs' analytical capabilities, or even the exploitation of strategies that human traders could not even conceive.[153] Moreover, the speed at which these systems operate, together with the ways algorithms interact with each other, can not only create contagion effects but may further complicate any foreseeability legal test.

Overall, autonomous AI is inherently prone to jeopardizing and undermining established prohibitions of market abuse, specifically traditional liability rules. Enforcement authorities are called upon to assess liability among a long list of possible individuals, but this can only be possible through knowing their exact contributions to

---

151 *See* Bathaee *supra* note 84 at 922-25 (discussing the interplay between the safe application of the "foreseeability" legal concept and the "black box" nature of specific AI methods).

152 *See, e.g.*, Yadav, *supra* note 134, at 1077-82 (discussing the difficulties for prosecutors in relying on the "negligence" test, especially in enforcing against HFT strategies).

153 *See* Bathaee, *supra* note 84, at 924.

the AI misconduct.[154] In this context, the "black box" nature of specific AI applications can easily bypass existing legal frameworks.[155] "Black box" AI trading could ultimately render the global financial system vulnerable to malicious actors "externalizing" to other market participants the costs of their misconducts.[156] We should note, however, that the above legal concerns are not entirely new. Algorithms, including deterministic ones from years or decades ago, could already act independently and behave in unpredictable ways, especially when interacting and competing with rivals' algorithms.[157] However, truly autonomous AI adds a further layer of complexity. To the extent that autonomous AI is a "black box," the safe application of liability rules against market abuse can be severely impaired. Legal concepts such as "intent" and "causation" will most of the time fail to provide a sound conceptual legal framework for authorities to enforce market conduct rules. Distinguishing whether AI misconduct results from an unintended consequence, inspired by some prior human intent, or autonomous AI decision-making will be increasingly challenging. As such, the matter has to be put on the interdisciplinary research agenda bridging financial law, economics, and informatics.

c. An Already "Outdated" Regime of Algorithmic Governance?

Global regulators and supervisors monitor carefully market-driven innovation in AI trading to assess the need to upgrade their legal frameworks. Notably, the governance of algorithmic trading is subject to some *lex specialis* in most advanced jurisdictions.[158] However, in dealing with specific AI trading methods, even the

154 AI is always bound to algorithms, which are coded to work as software on some sort of hardware (from desktops up to platforms in the cloud).

155 *E.g.*, Bathaee, *supra* note 84, at 919; *see also supra* note 56.

156 *See, e.g.*, Yadav, *supra* note 134, at 1039, 1083.

157 *E.g.*, Wendell Wallach, *Implementing Moral Decision Making Faculties in Computers and Robots*, 22 AI & SOC'Y 463, 464 (2008) ("Computers already operate independent of direct human supervision and make decisions that can't be predicated by their designers or programmers.").

158 The present section mainly builds on the EU financial law as an illustrative example, as being considered the most comprehensive legal system on algorithmic trading and its governance. In the U.S., algorithmic trading is regulated by both the Securities Exchange Commission and the Financial Industry Regulatory Authority (FINRA), a self-regulatory organisation intended to regulate member investment firms and exchange markets. *See* FINRA RULE 3110 and RULE 3120.

most advanced legal systems appear somewhat "outdated" as the technology evolves. Both the regulatory framework for algorithmic trading and market abuse enforcement mechanisms have some shortcomings.[159]

First, firms using algorithmic trading need to comply with specific requirements.[160] Importantly, they need to notify their algorithmic systems and strategies to both trading venues and competent authorities, and, upon request, provide to the latter information about their trading systems and controls,[161] with additional burdens for those firms conducting HFT or market-making activities.[162] Whenever those strategies involve increasingly autonomous AI trading that constitutes a "black box," it is uncertain whether supervisors enjoy the knowledge and expertise necessary for the proper oversight of algorithmic trading. Moreover, the law prescribes a set of organizational requirements to assist firms in compliance with market conduct rules, emphasizing the critical scope of enterprise risk management. Among those, focus should be placed on specific provisions regarding the "testing," "validation," and "deployment" of algorithmic trading strategies.[163] However, as the *European Securities and Markets Authority* (ESMA) recently highlighted in its revision report on the governance of algorithmic trading in the EU markets, regulators still rely upon the annual self-assessment report filled by the same supervisees as a base to ascertain their compliance. Now, it can be doubtful as to whether this mere state of compliance can really accommodate the increasingly sophisticated and "black box" nature of specific ML methods for financial trading (e.g., DRL). In this regard, the ESMA is considering the policy option to introduce a real due diligence

---

159 *But see* Peter Georg Picht & Gaspare Tazio Loderer, *Framing Algorithms: Competition Law and (Other) Regulatory Tools*, 42 WORLD COMPETITION 391 (2019) (arguing that financial regulation is one area of law that have successfully implemented rules and procedures to deal with issues arising from algorithms).

160 *Id.* at 395.

161 For the EU case, see Council Directive 2014/65/EU, art. 17 para. 2 of May 15, 2014, on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU, O.J. (L173/349) [hereinafter MiFID II].

162 For the EU case, see MiFID II, *supra* note 161, art. 16 para. 6 (general algorithmic trading requirements), art. 17 para. 2 subpara. 5 (HFT-specific requirements), art. 48 para. 10 (flagging of algorithms).

163 In the EU, these provisions are contained in Level 2 regulation. *See* Commission Delegated Regulation (EU) 2017/589 of 19 July 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council with regard to regulatory technical standards specifying the organizational requirements of investment firms engage in algorithmic trading.

process, not least to address many of these technical challenges regarding compliance.[164] Next, to mitigate market abuse risks, investment firms are expected to invest in some precautionary measures to counter the occurrence of unintended outcomes. Those remedies mainly include investments in internal systems and controls to monitor and mitigate risks from algorithmic trading.[165] But those can be hard to implement effectively, as they are notoriously costly and require high-level human expertise.[166] Thus, it is doubtful whether firms employing autonomous AI trading would always take all the necessary precautionary steps to be compliant.[167] Malicious market actors may be incentivized to use or discover abusive AI trading strategies, especially if these can assure them significant "alpha" (i.e., excess return due to the strategy) and do not expose firms to unaffordable legal and reputational risks.

Second, trading venues hosting algorithmic trading have some legal requirements to fulfill, including arrangements on trading systems' operational resilience,[168] circuit-breakers to moderate extreme volatility,[169] and electronic trading.[170] In this last regard, venues operators are called upon to cooperate with investment firms on aspects concerning algorithmic trading's conformity to market conduct rules, by providing, for instance, simulation environments to test algorithmic strategies.[171] Arguably, the fact that both investment firms and trading venues are required to prove

---

164 *Cf.* EUROPEAN SEC. & MKTS. AUTH., CONSULTATION PAPER: MIFID II/MIFIR REVIEW REPORT ON ALGORITHMIC TRADING 40 (2020) [hereinafter ESMA].

165 For the EU, see MiFID II, *supra* note 161, art. 17 para. 1. For the U.S., *see* FINRA RULES 3110.

166 *See* Dirk A. Zetzsche, Douglas Arner, Ross Buckley & Brian W. Tang, *Artificial Intelligence in Finance: Putting the Human in the Loop*, 43 SYDNEY L. REV. 43 (2021).

167 *Cf.* FINRA, Regulatory Notice 15-09 on Effective Supervision and Control Practices for Firms Engaging in Algorithmic Trading Strategies (Mar. 2015), https://www.finra.org/rules-guidance/notices/15-09 [https://perma.cc/U5W6-XKMZ] ("[I]n addition to specific requirements imposed on trading activity, firms have a fundamental obligation generally to supervise their trading activity to ensure that the activity does not violate any applicable FINRA rule, provision of the federal securities laws or any rule thereunder.").

168 For the EU case, see MiFID II, *supra* note 161, art. 48 para. 4.

169 For the EU case, see MiFID II, *supra* note 161, art. 48 para. 5.

170 For the EU case, see MiFID II, *supra* note 161, art. 48 paras. 6-10.

171 *See* Commission Delegated Regulation (EU) 2017/584 of 14 July 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council with regard to regulatory technical standards specifying organizational requirements of trading venues.

compliance through a self-assessment report does not seem entirely appropriate to deal with specific AI trading strategies of an autonomous nature.[172] Besides, in the oversight of market conduct rules, trading venues assume delegated supervisory tasks and need to have systems in place for market surveillance.[173] However, as watchdogs, they can face some incentive dilemmas. For obvious commercial reasons, they may not be sufficiently incentivized to conduct rigid screening on algorithms and market surveillance, especially when facing highly competitive pressure from other venues seeking to attract customers.[174] Furthermore, the possibility of an algorithmic trading firm being the operator of a trading venue (e.g., a dark pool operator) may give rise to conflict-of-interest concerns,[175] which could affect market surveillance. Another critical limitation of the current supervisory architecture is its limited scope for cross-market surveillance, which can indeed be a source of supervisory failure.[176]

Finally, apart from some essential regulatory competences, market conduct authorities usually have a relatively marginal role in market surveillance. In the enforcement of market rules, they generally rely on close collaboration with regulated market participants, from which they receive warnings and information about possible infringements.[177] With increasing market and

---

[172] *Cf.* FINRA Rules 3110 (Supervision); *cf. also* ESMA, *supra* note 164, at 44 paras. 142-45.

[173] *See* Janet Austin, *Unusual Trade or Market Manipulation? How Market Abuse is Detected by Securities Regulators, Trading Venues and Self-Regulatory Organizations*, 1 J. FIN. REG. 263, 266-74 (2015) (discussing market surveillance arrangements in some of the most advanced jurisdictions).

[174] *See* Yesha Yadav, *Oversight Failure in Securities Markets*, 104 CORNELL L. REV. 101, 104, 130-43 (2019) (arguing that market fragmentation and high competition between trading venues represent a barrier to effective oversight of market conduct rules); *see also* Austin, *supra* note 135, at 34 (discussing how the privatization of stock exchanges has led to greater conflicts of interest challenges).

[175] *See* Danny Busch, *MiFID II: Regulating High Frequency Trading, Other Forms of Algorithmic Trading and Direct Electronic Market Access*, 10 L. & FIN. MKTS. REV. 72, 75 (2016) (commenting the risk on the EU markets); *see also* Stanislav Dolgopolov, *Legal Liability for Fraud in the Evolving Architecture of Securities Markets*, *in* GLOBAL ALGORITHMIC CAPITAL MARKETS: HIGH FREQUENCY TRADING, DARK POOLS, AND REGULATORY CHALLENGES 272-73 (Walter Mattli ed., 2019) (providing evidence from a U.S. perspective).

[176] *See supra* note 134; *see also* ESMA, *supra* note 63, at 128-33 (assessing the need to establish a centralized cross-market surveillance mechanism at the EU level).

[177] *See* Yadav, *supra* note 174 (discussing the supervisory architecture in the U.S.); *see also* Busch, *supra* note 175, at 79 (for a European perspective).

regulatory fragmentation, malicious actors can find it easier to camouflage their abusive trading by hiding it in complex strategies within highly networked global markets.

Overall, it is somewhat questionable whether existing rules on algorithms and their governance can cope with increasingly autonomous AI trading agents' specificities. Tensions in achieving real transparency and accountability for specific ML methods (e.g., deep learning) may arise. At least, *de jure*, a strong form of explainability for algorithmic trading systems is required to comply with financial laws.[178] Importantly, however, compliance with "strong" explainability requirements creates a trade-off between ML models' utmost level of accuracy and possibilities of explanation.[179] In other words, the "black box" nature of specific ML methods (e.g., deep learning) applied to financial trading can frustrate firms' ability to comply with financial laws, but this also hampers the safe and legal implementation of specific AI trading strategies. However, it should be observed that, *de facto,* current legal systems and supervisory arrangements address "strong" AI explainability only in a limited way (e.g., most compliance exercises rely on self-assessment reports by investment firms and trading venues). In this context, regulators may face a dangerous trade-off, finding an optimal balance between technological neutrality and market integrity.

d. Keep It Closed, but "Fair-ly" White!

The apparently unsurmountable legal problems discussed above prompt the question of whether the existing liability rules need upgrading in view of greater AI autonomy and the "black box" problem. Over recent years, a rich discussion in scholarship and policymaking worldwide has emerged on a range of legal principles fit to cope with truly autonomous algorithmic decision-making. For instance, some commentators propose granting legal personality to autonomous AI agents, thus rendering the latter directly

---

178 *See* Adrien Bibal, Michael Lognoul, Alexandre de Streel & Benoît Frénay, *Legal Requirements on Explainability in Machine Learning*, ARTIFICIAL INTEL. & L. 16 (2020) ("[A]s total understanding of the model is required . . . [I]n the case of financial algorithms.").

179 *Id.*

accountable for their misbehavior.[180] Moreover, endowing AI with legal personality could be complemented by *ad-hoc* insurance coverage.[181] On the positive side, such a move would hopefully empower markets to "internalize" the costs of regulating AI. On the downside, it could increase moral hazard and even expose markets to new sources of systemic risks, not to mention all the difficulties for insurance companies in pricing new and emerging financial risks.

More extreme approaches would suggest imposing a strict ban on AI activities, whenever the risks at stake for society outweigh the related benefits.[182] Others have advocated applying a "strict liability" rule under tort law for harm caused by AI.[183] However, both proposals do not seem to fit well with the purposes and rationales of financial market regulation, inspired by principles of economic freedom, competition, and technological neutrality.[184] Both could indeed impair innovation, thus losing out on several potential efficiency gains.

It is true that liability rules will likely be unable to deal with the "black box" nature of specific autonomous AI methods. Nevertheless, they may still suffice when complemented by sound regulation addressing algorithmic trading governance. On this last point, indeed, the challenges brought by AI application in capital markets have prompted a lively debate on how to leverage the role of regulation to curb excess risk-taking and moral hazard, while stimulating innovation at the same time. Many interesting ideas

---

180 *See, e.g.*, JACOB TURNER, ROBOT RULES: REGULATING ARTIFICIAL INTELLIGENCE 185-86, 197 (2019) (arguing however that there might be the need to set some minimum criteria for AI personality).

181 *Cf.* Zetzsche et al., *supra* note 166, at 35.

182 *See* EUROPEAN COMMISSION, WHITE PAPER ON ARTIFICIAL INTELLIGENCE—A EUROPEAN APPROACH TO EXCELLENCE AND TRUST 10 (Feb. 19, 2020) [hereinafter COMMISSION WHITE PAPER], https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf [https://perma.cc/PB86-KV24] (quoting the German Data Ethics Commission's five-level risk-based regulation system, which envisages a complete ban for the most dangerous AI applications).

183 *See, e.g.*, Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61 (2019) (arguing that traditional products liability rules could be applied to certain AI applications); Bathaee, *supra* note 84, at 931-32 (discussing the trade-off between safety and innovation in imposing strict liability rules).

184 *See, e.g.*, Pedro Magalhães Batista & Wolf-Georg Ringe, *Dynamism in Financial Market Regulation: Harnessing Regulatory and Supervisory Technologies*, 4 STAN. J. BLOCKCHAIN L. & POL'Y 203 (2021) (discussing the need for greater integration between regulatory technology and supervisory technology).

have emerged on how to deal with the specificities and additional risks of AI applications; evaluating them in detail is, however, beyond the scope of this study. Nevertheless, in what follows, we discuss a number of guiding principles for the effective implementation and use of AI applications, with a view to initiating a critical policy discourse on this matter.

One frequently debated regulatory approach, which also leverages existing internal governance frameworks, concerns keeping a "human-in-the-loop" in all AI decisive processes, in order to guarantee personal responsibility and accountability. The main idea is that, by assigning specific roles and responsibility to individuals alongside the AI supply chain, "traceability" would alleviate many liability attribution challenges.[185] Positively, this regulatory option will also entail strengthening current legal frameworks, without significant and radical law revisions, by emphasizing AI processes' transparency and auditability. Importantly, the "the-human-in-the-loop" approach is also supported by several recent policy initiatives worldwide. Indeed, it has been adopted by some "soft law" instruments aimed at guiding private organizations towards trustworthy implementations of AI,[186] especially when those imply high-risk decisions.[187] In the

[185] *E.g.*, Zetzsche et al., *supra* note 166, at 38-39, 46-48 (discussing however all the practical challenges of promoting such an approach).

[186] *E.g.*, High-Level Expert Group on Artificial Intelligence, *supra* note 38, at 14-20 (proposing seven requirements for trustworthy implementations of AI, namely: (i) human agency and oversight; (ii) technical robustness and safety; (iii) privacy and data governance; (iv) transparency; (v) diversity, non-discrimination and fairness; (vi) societal and environmental wellbeing; (vii) accountability).

[187] *E.g.*, COMMISSION WHITE PAPER, *supra* note 182, at 18-22 (proposing six types of requirements for high-risk applications of AI, namely: (i) training data; (ii) data and record-keeping; (iii) information to be provided; (iv) robustness and accuracy; (v) human oversight; (vi) specific requirements for certain particular AI applications).

same vein, financial authorities,[188] self-regulated,[189] and private organizations alike,[190] are all engaged in fostering the AI regulatory science. Most proposals go into the direction of an enhanced "precautionary" approach to AI regulation, requiring increased coordination among different stakeholders. Notably, public authorities aim at ensuring trustworthy AI development and implementation by affirming the relevance of well-designed model risk management, data governance, and compliance requirements as the most fundamental AI principles and high regulatory priorities. In a nutshell, policymakers' best bet is to create and promote best practices within the industry to incentivize market participants to effectively take due care of their algorithms.

In the area of algorithmic market abuse, we have argued that explaining algorithms' outcomes and behaviors is a crucial element in ascertaining and attributing liability for AI wrongdoing. Indeed, the "black box" nature of specific ML methods that allows for autonomous trading agents (e.g., DRL) could be interpreted as a transparency problem. Innovative ideas on how to deal with such an issue suggest moving from typical "command-and-control" to embracing more dynamic and flexible approaches, through a combination of both *ex-ante* and *ex-post* regulation, leveraging the

---

188 In Singapore, the Monetary Authority of Singapore was among the first authorities to provide a special framework for the development of AI applications in the financial services industry. *See* MONETARY AUTH. OF SINGAPORE, PRINCIPLES TO PROMOTE FAIRNESS, ETHICS, ACCOUNTABILITY AND TRANSPARENCY (FEAT) IN THE USE OF ARTIFICIAL INTELLIGENCE AND DATA ANALYTICS IN SINGAPORE'S FINANCIAL SECTOR (2018), https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf [https://perma.cc/3UMS-FUZ2]. For an overview of recent regulatory efforts by financial authorities worldwide, see Zetzsche et al., *supra* note 166, at 28-34 (reporting on the cases of the European ESAs, De Nederlandsche Bank, and the Hong Kong Monetary Authority).

189 *See, e.g.*, FINRA, ARTIFICIAL INTELLIGENCE (AI) IN THE SECURITY INDUSTRY (2020), https://www.finra.org/sites/default/files/2020-06/ai-report-061020.pdf [https://perma.cc/RJZ6-FJRL] (highlighting the main factors firms when seeking to adopt AI applications need to consider, namely: (i) model risk management; (ii) data governance; (iii) customer privacy; (iv) supervisory control systems; (v) cybersecurity; (vi) outsourcing and vendor management; (vii) record keeping; and (viii) workforce structure).

190 MICROSOFT, DEUTSCHE BANK, LINKLATERS, STANDARD CHARTERED & VISA, FROM PRINCIPLES TO PRACTICE: USE CASES FOR IMPLEMENTING RESPONSIBLE AI IN FINANCIAL SERVICES (2019), https://aka.ms/fromprinciplestopractice [https://perma.cc/82QJ-PRBA] (implementing Singapore MAS's FEAT principles for AI).

use of regulatory technology to face the new challenges brought by the evolving complexities of financial technology.[191]

From an *ex-ante* perspective, regulators can indeed explore a wide spectrum of policy options to enhance AI transparency. To start with, they would need to make important decisions, such as the scope of codes' inspection, and to whom supervisory responsibility should be attributed. While opening an AI "black box" to inspect model transparency can provide useful information about a code and its parameters that bridge input with output, this nevertheless works on the assumption that regulators enjoy some level of specific domain knowledge. But this is highly debatable.[192] On their part, investment firms will be most likely reluctant to disclose details about their proprietary projects, as they may rightly fear risks of some leakages to competitors giving rise to IPR and competition law concerns.[193] Opening the "black box" is *per se* a somewhat problematic policy option, as it can ultimately hamper trust, innovation, and competition.

Instead of aiming for transparency of the models, alternative views suggest testing AI trading strategies for their possible abusive tendency under different market conditions before allowing them to operate in the markets. For instance, this can be done by institutionalizing more regulated testing frameworks, which are already contemplated by existing financial laws, as part of a novel authorization regime for AI.[194] Whenever financial authorities are not well-positioned for such responsibility, the task could be delegated to a newly established independent third-party organization, which could arguably better represent and balance all stakeholders' legitimate interests.[195] Nevertheless, this policy option also requires thoughtful considerations regarding the scope of testing activities, since authorization to use specific AI trading strategies will be granted as an outcome of testing procedures. In other words, following authorization, a particular AI trading

---

[191] *E.g.*, Hilary J. Allen, *Driverless Finance*, 10 HARV. BUS. L. REV. 157, 195-96 (2020); Gina-Gail S. Fletcher, *Macroeconomic Consequences of Market Manipulation*, 83 L. & CONTEMP. PROBS. 123, 138-40 (2020).

[192] *Cf.* Allen, *supra* note 191, at 198-99; Zetzsche et al., *supra* note 166, at 48-49.

[193] Iain Sheridan, *MiFID II in the Context of Financial Technology and Regulatory Technology*, 12 CAP'L MKTS. L.J. 417, 420 (2017).

[194] *E.g.*, Allen, *supra* note 191, at 196-98.

[195] *See* Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83 (2017) (arguing in favor of the creation of a centralized regulatory agency for the governance of algorithms).

strategy will be certified against known forms of market abuse (i.e., "good conduct" by design). However, regulators would need to make important decisions regarding the regulatory framework for the development of ML methods, including specific arrangements regulating the role of all individuals involved in the AI supply chain, as well as the role of training data to be used in simulated scenarios. In principle, authorization should be calibrated on a risk-based approach. Depending on the specific case, AI trading systems will need to be tested according to specific market structures and trading dynamics to assess the likelihood of them being disruptive. Still, questions remain on how the tendency of AI to misbehave can be ascertained, especially for new risks of market manipulation (e.g., "aggressive" cross-asset and cross-market trading) and "tacit" collusion. In fact, regulators would need a framework to clearly distinguish legitimate trading from unlawful strategies, which instead have the sole purpose of putting prices under pressure or triggering other market participants to the same effects without any justified economic interest.[196] Alternatively, for the risks of "tacit" collusion, a framework would need to determine when an AI trading agent behaves in a non-competitive manner.[197] Ultimately, the effectiveness of any pre-approval regime through testing is based on the latter's ability to uncover AI ability to learn how to game market rules in simulated environments. Yet there can arguably be gaps between simulated and real-market environments. On a positive note, market authorities are themselves engaged in technological innovation to enhance their oversight capabilities (i.e., supervisory technology). Using AI to monitor AI seems indeed a necessary step to achieve effective supervision.[198] However, effective supervision and auditability of AI hinge on assumptions and training data being available for testing purposes. In particular,

---

196 *See, e.g.*, David C. Donald, *Regulating Market Manipulation Through an Understanding of Price Creation*, 6 NTU L. REV. 55 (2011) (arguing that to regulate market manipulation effectively, regulators need first a clear and proper understanding of markets and price creation mechanism); *see also* Matthijs Nelemans, *Redefining Trade-Based Market Manipulation*, 42 VAL. U.L. REV. 1169, 1183 (2008) (arguing that regulators should tackle those trading strategies causing "unsupported price pressure").

197 *See generally* Joseph E. Harrington, *Developing Competition Law for Collusion by Autonomous Artificial Agents*, 14 J. COMPETITION L. & ECON. 331, 356-58 (2018) (developing a three steps framework to determine the lawfulness of algorithms).

198 Lawrence G. Baxter, *Adaptive Financial Regulation and RegTech: A Concept Article on Realistic Protection for Victims of Bank Failures*, 66 DUKE L.J. 567, 600-03 (2016); Allen, *supra* note 191, at 203-05.

whenever simulated scenarios differ from real use cases, the same testing procedure may lose significance and have no power to detect risky AI trading strategies.

Furthermore, *ex-post* regulatory options can help to ensure *ex-ante* measures' effectiveness and, more generally, AI trading regulation.[199] *Ex-post* regulatory options should be considered to enhance AI auditability, not least to address certain self-learning AI trading methods' somewhat kaleidoscopic behavior. After all, regulators and supervisors need to audit algorithms' behaviors and their potential harm with respect to market integrity. When looking at current supervisory mechanisms and infrastructures in place, however, there are a few reasons to believe that they were conceptualized for older times, which were characterized by lower market fragmentation and the presence of less autonomous algorithms. To remedy this, recent findings from the field of regulatory technology ("RegTech") and supervisory technology ("SupTech") would suggest leveraging the role of the technology itself.[200] In highly fragmented, super-fast trading, and mainly algorithmic global capital markets, there is indeed a need to rethink our global supervisory architecture to deal with the challenges brought by algorithmic market abuse, one that ensures cross-market surveillance. To this end, however, it is doubtful whether trading venues alone can take on this task, or whether novel public-private partnerships would be desirable.[201] Arguably, market conduct authorities could have a more significant role in market surveillance *vis-à-vis* private organizations. The latter, because of their legitimate business interests, could indeed compromise market integrity in the

---

[199] Allen, *supra* note 191, at 203.

[200] *See, e.g.*, FINRA, TECHNOLOGY BASED INNOVATIONS FOR REGULATORY COMPLIANCE ("REGTECH") IN THE SECURITIES INDUSTRY 2 (Sept. 2018), https://www.finra.org/sites/default/files/2018_RegTech_Report.pdf [https://perma.cc/66MM-LQLL] ("[M]arket participants are increasingly looking to use RegTech tools to help them develop more effective, efficient, and risk-based compliance programs"); DIRK BROEDERS & JERMY PRENIO, BANK OF INT'L SETTLEMENT, INNOVATIVE TECHNOLOGY IN FINANCIAL SUPERVISION (SUPTECH)—THE EXPERIENCE OF EARLY USERS (2018), https://www.bis.org/fsi/publ/insights9.pdf [https://perma.cc/NW89-HZ5E] (surveying the use of SupTech by a number of supervisory authority among most advanced jurisdictions).

[201] *See, e.g.*, Yueh-Ping (Alex) Yang & Cheng-Yun Tsang, *RegTech and the New Era of Financial Regulators: Envisaging More Public-Private-Partnership Models of Financial Regulators*, 21 U. PA. J. BUS. L. 354 (2018) (arguing that financial regulators would benefit from an enhanced public-private partnership and discussing four different models for such a collaboration including (i) mixed-ownership RegTech organization, (ii) contracted RegTech supporter, (iii) a quasi-public RegTech regulator, and (iv) directly delegated gatekeepers).

name of private profits. However, new innovative solutions may also emerge from the market and further shape the competitive landscape of the market surveillance business.[202] Moreover, surveillance mechanisms should also be improved to allow for "real-time" market conduct supervision,[203] a solution that would most likely ease enforcement actions against algorithmic market abuse, but that simultaneously highlights the fundamental importance of coordination among different stakeholders.[204] To complement this, reporting arrangements could be strengthened to provide supervisors with timely information regarding specific trading strategies used by market participants that employ sophisticated AI trading systems.[205] Lastly, RegTech could also bridge supervisors and supervisees more closely and allow for constant regulatory dialectic. For instance, some emerging initiatives have explored the merits and feasibility of introducing some forms of machine-readable regulation endorsement to upgrade current regulatory tools.[206] By doing so, it is expected that financial authorities could directly deal with AI trading systems as they go on markets, without the need for them to always mediate with AI users.

---

[202] *See, e.g.*, Holly A. Bell, *Chapter 10: Using the Market to Manage Proprietary Algorithmic Trading*, *in* REFRAMING FINANCIAL REGULATION: ENHANCING STABILITY AND PROTECTING CONSUMERS 266-68 (Hester Piece & Benjamin Klutsey eds., 2016) (noting the relevance of cooperative market-based solutions to minimize competition between regulators and market participants on the development of market structure and surveillance mechanisms).

[203] *See* FINRA, *supra* note 200, at 4 ("[T]raditional rule-based systems to a predictive, risk-based surveillance model that identifies and exploits patterns in data to inform decision-making.").

[204] *See, e.g.*, Douglas W. Arner, Janos Barberis & Ross P. Buckley, *FinTech, RegTech and the Reconceptualization of Financial Regulation*, 37 NW. J. INT'L L. & BUS. 371 (2017) (discussing the potential of "regulatory technology" to enable a close to real-time and proportionate regulatory regime to balance expected risks and efficient compliance, also for the case of market manipulation).

[205] *E.g.*, Gina-Gail S. Fletcher, *Legitimate Yet Manipulative: The Conundrum of Open-Market Manipulation*, 68 DUKE L.J. 479, 542-43 (2018).

[206] *See* FINRA, *supra* note 189, at 9 (confirming that regulators are exploring and adopting the concept of machine "machine-readable" rulebooks, which could arguably allow firms to automate regulatory compliance internal processes); *see also* Eva Micheler & Anna Whaley, *Regulatory Technology: Replacing Law with Computer Code*, 21 EUR. BUS. ORG. L. REV. 349, 362-64 (2020) (discussing possible barriers for the effective implementation of regulatory technology solutions to deliver machine-readable code onto existing IT systems, but also regulatory capture risks for the future development of regulatory technology projects); Schwalbe, *supra* note 98, at 599 (suggesting the idea to incorporate legal provisions and constraints into algorithms themselves, similar to the three Isaac Asimov's robotics laws).

To conclude, there is one main takeaway from all of this. When applying traditional regulatory approaches to the challenges brought by new financial technologies, the law frequently falls short. This is precisely the case for increasingly autonomous AI trading systems (i.e., DRL) and their implications for the reliable enforcement of market abuse regulations. Crucially, whenever policymakers attempt to achieve legal simplicity and market integrity, without hampering innovation, they have reached only two out of three policy goals at best.[207] As this paper has attempted to demonstrate, market conduct regulators face the same policy "trilemma" in approaching regulation to deal with increasingly autonomous AI trading strategies, their "black boxes," and new forms of algorithmic market abuse.

V. CONCLUSION

AI trading is an evolutionary step forward in algorithmic trading techniques. Continuous progress in ML methods applied to financial trading will pave the way for a new computational finance paradigm: as we have seen, "deep reinforcement learning" trading methods will allow for approximating (truly) autonomous AI trading systems probably evolving into nearly or even truly autonomous AI trading agents in the long run. While increasingly autonomous AI agents are proposed to deliver several efficiency gains for both organizations and markets, their AI agency raises fundamental ethical and legal questions of liability in cases of wrongdoing. Our exploratory study has approached these issues from a financial market conduct perspective. Through a number of illustrative examples, we have conceptually shown that autonomous AI trading methods will be able to allow for both old and new forms of market manipulation, including emerging risks of algorithmic "tacit" collusion. The above-discussed novel scenarios of market abuse by autonomous AI trading systems, primarily elicited by their "black box" nature, pinpoint a number of open questions. Importantly, established liability rules (e.g., "intent" and "causation") do not sufficiently cover instances of autonomous AI

---

[207] *See* Chris Brummer & Yesha Yadav, *FinTech and the Innovation Trilemma*, 107 GEO. L.J. 235 (2019) (arguing that to alleviate the trilemma's effects, regulators should enhance their institutional arrangements to achieve greater domestic cooperation and international coordination and rely on more self-regulation by market actors).

decision-making. Notably, whenever AI amounts to a "black box," liability attribution rules may be subverted. Moreover, existing enforcement mechanisms, including market surveillance arrangements, can become outdated as being increasingly unable to police those forms of market misconduct led by algorithmic agents.

In view of these regulatory shortcomings, we have discussed a number of policy proposals that have been put forward as to legal reform, and we develop several guiding principles to inform a sound policy response. While we remain skeptical as to the real-world effectiveness of any *ex-ante* screening mechanism, we would put greater hope in robust governance requirements and ongoing monitoring arrangements. For the near future, mandating a "human-in-the-loop" seems the only viable regulatory option, not least to foster a culture of responsible and safe AI development. Undoubtedly, as a society, we would prefer avoiding inadequate and negligent human oversight on risky activities.

Market abuse usually starts as a local phenomenon, a mere attack to a particular market's integrity. However, in a globalized economy, whenever safeguard mechanisms fail to contain the risks, algorithmic market abuse can also go viral and may spill over to the whole financial system to the point of threatening its systemic stability, thus testing the resiliency of our global economy at large. For all these reasons, the need to rethink our regulatory toolkit is more than urgent. As this study has attempted to show and promote, there is a dire need for a change in academic attitude in favor of more interdisciplinary research and education for better meeting the complex challenges of financial technology and innovation in order to develop synergies between the scientific fields of financial law, economics, and informatics.