AI Legal Counsel to train and regulate legally constrained Autonomous systems

Hal Ashton*
Computer Science
University College London
London, UK
https://orcid.org/0000-0002-1780-9127

***Abstract*—In this paper I motivate and present the idea of a legal counsel for autonomous algorithmic agents (or A-bots). In its simple incarnation the counsel is a legal oracle which is able to tell the A-bot the legality of any action at any moment in time. I enumerate a number of capabilities that the oracle requires to work in a regulated environment with a rich set of rules. A point of difference is that the oracle can be used for training and online control of A-bots and is relatively independent to the learning methodology chosen. It is an example of a hybrid system known as a shield, which uses methods from both symbolic and statistical AI.**

***Index Terms*—Autonomous systems, Learning (artificial intelligence), Decision support systems, Law, Product safety engineering, Software safety**

I. Introduction

Autonomous algorithmic agents (hence A-bots) are more commonly operating in regulated areas; as their autonomy increases so does their ability to cause harm.[1] To date, engineers have mostly concerned itself with scenarios where A-bots might realise physical harm to themselves or others. This is understandable in areas such as aerospace where plane autopilots have enormous responsibility. Legal constraints are fundamentally different in the sense that their transgression may not be obvious to the A-bot or its owner. Moreover, breaking the law may be advantageous to an A-bot's objective function whereas the breaking of physical constraints are typically not. Determining the legality of behaviour may require analysing long sequences of past behaviour, can depend on why actions were taken and may even be determined by planned actions never realised by the A-bot. In short, getting an A-bot to respect the rule of law is fundamentally harder than getting it not to crash into things which itself is far from trivial. To complicate matters further, the two tasks might, on occasion, contradict each other.

When we think of A-bots we are drawn instinctively to their physical manifestations such as Autonomous vehicles. However, whilst they are certainly exciting, they are things of the future. The A-bots that already exist and are widely deployed, mostly do not have bodies but nevertheless have the scope and opportunity to cause harm of the non-kinetic variety. Algorithmic trading in regulated financial markets has been a feature of the market for decades now, and has some negative effects (see Manahov [2]). In a more quotidian setting, likely to directly affect all of us, Calvano et al [3] show how price setting algorithms of the type that an e-commerce company might use, learn how to tacitly collude with each other. Putting aside the thorny question of liability in such a case (A-bots are not legal entities and not capable of committing crimes). In the simple price-setting case how would the A-bot's programmers deploy their A-bot to take advantage of the many wonderful properties that AI has whilst being comfortable that it did not break the law?

Just as companies employ legal counsel to advise them on issues of the law governing their conduct, I believe that there will be a demand for A-bots to use a legal-counsel in an online[2] manner. Whilst this legal counsel *could* be human, it seems likely that it will itself be AI driven for reasons described in the next section. Etzioni and Etzioni [4] call these secondary AI Monitor systems *Guardians*. Note that the AI counsel is capable of fulfilling two functions - firstly as a training prior to deployment it will teach the A-bot legal behaviour implicitly. This is desirable because it means machine learning methods can be used to allow A-bots to learn novel and efficient ways of completing their tasks. Secondly in deployment it will regulate the A-bot's behaviour. This is important in settings where the chance of unexpected situations is high regardless of the time spent training.

II. Motivating an AI counsel for A-bots

A legal counsel for A-bots is likely to be powered by AI for the following reasons:

- **Volume** Modern machine learning methods are notoriously wasteful with data when training in simulation environments, an AI legal counsel will need to respond to a lot of requests when training A-bots.
- **Speed** Many A-bots operate in time critical environments (like trading or driving in the future) where a short response time is a requirement.

*Supported by EPSRC, UK

[1]See Wellman and Rajan [1] for a specific example of this effect in trading algorithms within finance.

[2]Here online means 'within a sequential learning and acting process' and not to do with the internet sense of the word online

978-1-7281-6251-5/20/$31.00 ©2020 IEEE

- **Scaleability** Parallelisation is a powerful technique to speed up algorithm training where many versions of an A-bot are trained independently at the same using different processor cores. An AI legal counsel would be available to multiple A-bots at any time. After deployment, this means that AI counsels are able to look after multiple clients concurrently. This will allow a legality software as a service industry.
- **Triviality** Most requests are likely to be trivial in their nature as Prakken notes [5]. Simple requests are better served by (free) AI than expensive legal minds.
- **Transportability** A marketable A-bot is likely to require the ability to work in many different countries. No individual law firm may have the ability to service the required geographic footprint.

In the first instance it is probable that the AI counsel would work as a legal oracle, giving the A-bot a range of actions that it could take immediately depending on the situation. This reduces the type of question that the A-bot receives to one, simplifying its design and fast forwarding its deployment ahead of the AGI powered legal counsel required for more unstructured questions. The format of response sent to the A-bot is just the set of admissible actions that it can legally perform next. The design is shown in Fig. 1. The A-bot receives information about the state of the world it is in, it then queries the Oracle as to what actions it can legally take next. The Oracle will examine the A-bot's information (history of action and intentions) and refer to its explicit Rule book. It will decide which actions are admissible or not and convey the information back to the A-bot. The A-bot will then choose the best action according to its objective function. The world will then evolve according to the A-bot's action and other factors and if the A-bot is training itself through Reinforcement Learning, it will receive a reward and the cycle repeats.

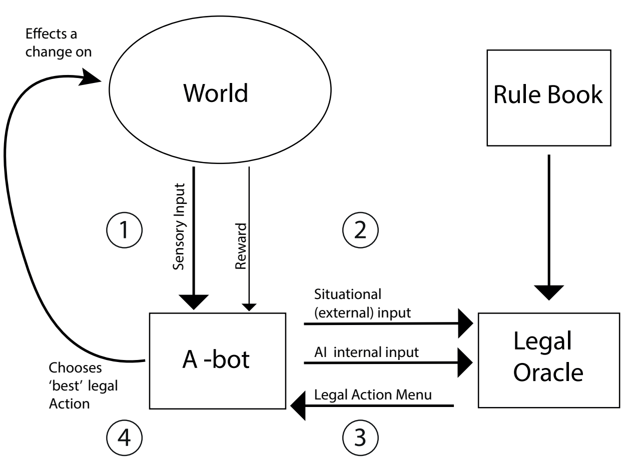

Fig. 1. 1. A-bot receives sensory input 2. Queries Legal oracle about possible actions 3. Receives list of legal actions 4. A-bot chooses best action according to its objectives

III. Required Legal Oracle Capabilities

If the legal oracle is to succeed in its task of showing which actions are legal at any point in time, then it needs to overcome a number of obstacles which I will summarise as follows:

1) **Rule book**: The Oracle requires a comprehensive rule book which describes in a machine readable format, what the A-bot is and is not allowed to do.
2) **Sufficient Situational Information**: The Oracle requires sufficient historic information to work which may be greater than which the A-bot uses to make decisions.
3) **Predictive Causality**: The Legal Oracle needs to look forward to ascertain the likely causal effect of each action in order to determine their legality.
4) **Intent**: The legality of certain behaviours rests on the intent behind those behaviours. The Legal Oracle needs to determine why an A-bot is doing something before judging whether that is allowable or not.
5) **Determination**: Armed with the information in 2 to 4 the oracle needs to have the ability to assess whether any action will break the rules contained in 1.

In the following subsections I will discuss each of the enumerated requirements in greater detail.

A. Rule book

The Oracle must have access to a rule book which is comprehensive and specific to the environment that the A-bot is operating in. How laws are described as code is a separate subject beyond the scope of this paper. Some progress is being made; Alves et al [6] successfully encode road junction rules using temporal logic. Whilst I refer to 'a Rule Book', in practice it will almost certainly not be a single source since depending on where the A-bot is and what it is doing, laws will almost certainly be not be completely defined in a single statute book. Instead the Rule Book would have to be created from any number of relevant sources (like case law). See Boella et al [7] for a more lengthy criticism of the textualist approach to law adopted by computer scientists. One observation is that within any ruleset, there are likely to be contradictory laws. The Rule Book that the Oracle has access to needs to describe the hierarchy of laws so that in the event of stalemate or inescapable law breaking, the Oracle can advise. This is mentioned in the IEEE global initiative on Ethics of Autonomous systems albeit in the context of norms [8]. Furthermore in the case of Cyber-Physical Systems (CPS) considerations of legality will have to be balanced with those of safety.

B. Sufficient situational information inputs to the oracle

A wide variety of laws exist in the real world whose determination often requires a wider variety of information than is used by the A-bot to make decisions. For example, it is often enough for A-bots to assume a Markovian world where the current state of the world is sufficient information to estimate the transition of the world into another state given any action. However many laws are conduct specific (consider for example Reckless driving in the UK), which means that determining whether they are broken or not depends on the past trajectory of actions taken. This means that the legal oracle will require to receive or record the history of actions taken by the A-bot.

The requirement for the oracle to examine the history of world to determine the legal action set at any moment in time does not necessarily impact the Markovian assumption of the A-bot and the statistical properties that this imbues for learning. The oracle is in effect external to the A-bot and is as far as the the A-bot is concerned part of the environment. Consider the canonical example of the maze solving robot. When it is positioned on the boundary of the maze, actions which would otherwise move it in a direction leading to a position outside the maze are typically masked or given a null-effect. This is an unacknowledged Oracle which is telling the robot about the rules of the maze: "Don't bother trying to move left when you are on the leftmost position in the maze".

C. Predictive Causality

In civil law the establishment of causality is a key component when determining the liability of damage caused by an actor[3]. In criminal law its proof is similarly key to establishing whether a crime has been committed. Since the oracle will have to answer questions of the type: Will action a likely cause prohibited consequence k, the oracle must possess an adequate predictive causal model of the world.

Many notable achievements in Reinforcement Learning such as the Atari game players of Mnih et al [10] have come about in designs where the A-bot has no model of the world (in the language of RL, they are *model-free*). They therefore are not able to reason about the causal effects of their future actions, nor do they plan ahead in a sense that we humans understand the word.

Any accurate predictive causal model of the world can be used to answer the type of causal question that the Oracle requires since a sufficient answer requires $P(k|a)$ and $\sum_{b\in\mathcal{A}} P(k|b)$ for prohibited consequence k and action a in A-bot's action set $\mathcal{A}$. Halpern [11] calls the evaluation of future causality 'type causality' in contrast with his 'actual causality' which refers to the analysis of events passed and is built upon the idea of the counterfactual. It is an open question whether judgements of causality ex-ante align perfectly with judgements of causality ex-post. To be conservative it is highly desirable that an A-bot which judges itself not to be causing something undesirable before acting is not subsequently judged to have been a cause in the legal/evidential sense afterwards.

D. Intent

The legality of certain actions is sometimes defined by the intent behind them or to rephrase, the admissibility of an action might depend on the A-bot's planned future actions. An example of this in financial markets is within limit order book where the placement of orders with the intent to cancel and thereby mislead the market as to supply and demand is an offence known as spoofing [12]. Bathaee [13] calls this Basis Intent, noting that anti-discrimination laws in the USA also rely on it. Encroaching on the thorny topic of A-bot legal personhood[4], many crimes are crimes of specific intent - they can only be committed if a certain level of intent on the part of the defendent can be proved. Furthermore inchoate offences exists (intent to do X) which refer to future actions that might not be realised. A-bots will typically have a policy function Π which is a function from known situational information to action space. Given any situation, it will inform the A-bot how to behave. The Oracle will need access to this policy function and apply it to its model of the world mentioned in the previous section to check to see if the A-bot has intent over a future state of the world which it should not. It should be noted that the concept of intent is undefined formally for A-bots (see Ashton [15]).

E. Determination

Equipped with sufficient situational information, a predictive causal model of the world and the policy function of the A-bot, the Oracle then needs a mechanism to verify for each possible action within the A-bot's action set, whether it will cause a rule break according to the Rule Book. Though a large body of research exists on the subject with the research area of AI and law, as Prakken [16] observes the focus for the A-bot is for *future* actions, but existing research in the subject area of AI and laws has mostly concerned itself with determining the legality of actions already taken. Hasanbeig, Abate and Kroening [17] demonstrate a way of converting laws specified in temporal logic into an finite state automaton which is able to classify histories of actions as conforming or not. This technique, which originates in program verification research, is beginning to establish itself in Safe-RL though there have been no approaches specifically addressing legal behavioural restrictions.

IV. Discussion

This design of a legal oracle, separate from the A-bot itself, is termed a gateway system by the IEEE in [8] and a recommended design feature (albeit in the wider context of norm-embedding for A-bots) to provide a fail-safe for the behaviour of the A-bot. It is termed an *Ethical Governor* in Winfield, Blum and Liu [18] and a *Guardian* in Etzioni [4].

Where Reinforcement Learning (RL) is used as the learning mechanism, the Oracle should ensure that unlawful actions are never chosen during exploration. As mentioned, within RL there is precedent for constrained learning through the imposition of an external method - the A-bot which bested the World champion Go player [19] had access to a fully functioning game simulator where only legal moves were possible (it was never given the option of tipping the board over for instance!). To our knowledge, this has not been

[3]Excepting regimes of strict liability where causality is not necessary. See Turner [9] for a full exploration of liability regimes

[4]This is a large topic area, for an exploration see Chopra and White [14] or Turner chapter 5 [9]

explicitly acknowledged as a control mechanism for the learner in these contexts.

The Oracle structure presented in this paper is an example of a *shield*, which has been shown by Jansen et al [20] and Alshiekh et al [21] to work with RL to produce safe policies. Both approaches use types of Temporal logic to encode a set of laws and then constrain a learning algorithm to those laws during training and deployment. Techniques borrowed from formal verification applications (see [22]) are then used to ensure that the allowable actions taken do not contravene any laws. A risk when using deep RL, comes from the fact that the resulting policy function, which determines how the A-bot acts depending on its situation, is a function estimator of an optimal policy comprised of a multi-layer neural network. In the example of Go for example it would be impossible for a training process to consider all configurations of the board, so the policy function interpolates when it encounters novel situations. Because the policy function is a multi-layered neural network, this interpolation process is not interpretable. An A-bot trained to act legally may encounter an unusual situation in deployment which it had not met during training and there is no guarantee that its policy function would tell it to act legally. This is a problem which affects any type of exhaustive verification procedure in a complex system regardless of whether its design originates from statistical or symbolic AI [23]. In the architecture of Fig. 1, the legal oracle will also work in deployment so this risk is mitigated.

V. Related work

Prakken [16] observes that the existing research on the constraint of A-bots (specifically Autonomous Vehicles) to legal behaviour is limited. A problem which bedevils academic research in computer science is the existence of separate silos that research exists in (Boella et al [7] calls it 'fragmentation') and the ensuing lack of a common terminology between them. Even within a single subject area such as artificial intelligence, there is a notable divide between the symbolic AI and statistical (soft) AI research communities. This topic is no different. It is simultaneously called Value Alignment, Program Verification, The Norm Implementation Problem and Safe Machine Learning though much of the research in these areas refers to more general work on ethical and norm constraint not lawful constraint. On the subject of Value Alignment in behaviour, Arnold, Kasenberg and Scheutz [24] identify four areas of interest, namely Intentions, Reasons, Norms and Counterfactuals. A parallel can be drawn with the Oracle's requirements of Intent, A Rule book and Predictive Causality respectively.

The low quantity of research on methods to restrict A-bots to legal behaviour is in contrast with the growing body of research on the subject of constraining AI to behave ethically (for a review see Winfield [25]). Leenes and Lucivero [26] observe a gap between the ethical constraint debate and the more "mundane" questions that roboticists currently face, legality being one of those. On the subject of ethics and AVs, Lin [27] is correct in saying: *"Some decisions are more than just the mechanical application of traffic laws"*, but is optimistic about current capabilities in the application and comprehension of law within A-bots. As Theodorou and Dignum [28] state *Abidance to law is the basis of any application, whereas ethics is the 'sky'*[5] But what is the reason behind the lack of research on legal constraint vs ethical ones? Hildebrandt [30] posits that technologists can decide on ethics behind closed doors but laws are generally created with public oversight which is slow. Unlike Law, ethics cannot provide closure on any particular behaviour because it does not have the force of law, but Hildebrandt fears that technology rationalised with ethics will force closure. Cath [31] notes the high representation of AI industry members in policy forming and influencing organisations.

Existing methods to constrain A-bots can be divided between those that have an explicit rendering of laws to verify behaviour against and those that aim to learn them implicitly. In the latter case, a body of data is required to learn the rules of behaviour which they were generated under. Wallach and Allen [32] call the two approaches top-down and bottom up respectively. Whilst there is correlation with the use of symbolic AI in approaches where a ruleset is provided and statistical or soft AI where rules are deduced from data, they are not exclusively faithful, see Table I for a simple taxonomy. A system to derive symbolic rules from data could be used for example such as in association sequential rule mining. Likewise Safe Reinforcement Learning (RL) techniques (see Garcia and Fernandez [33] for a review) which are soft AI attempt to learn 'safe' policies typically subject to reward functions with penalties set according to a limited but explicit ruleset based on restricted states. A new class of soft AI learning techniques termed Seldonian has been recently proposed by Thomas et al [34] which ensure safe behaviour and admit more complicated rule representations. Charisi et al [23] explore the pros and cons of soft and symbolic approaches and make a good case that a hybrid approach, such as the one presented in this paper, can harness the desirable features of both.

Teaching A-bots ethical or legal behaviour by observing and training with large amounts of data is an approach associated with data intensive (soft) machine learning methods. An issue is that this training often involves a simulator else risky behaviour might endanger the A-bot. Simulators risk misspecification leading to unexpected behaviour being learnt (see Lehman et al [35]). One method is Apprenticeship learning which is a type of Inverse reinforcement learning (IRL) [36] where the policy function or reward function of an expert demonstrator is learnt by the A-bot though observation of their behaviour. Noothigattu et al [37] teach their A-bot to play Pac-man obeying the rule - don't eat the ghosts after eating a power pill - through expert demonstration. A bandit policy is then learned and used to efficiently switch between the safe policy demonstrated by the expert and an efficient

[5]The Phenomena of horizon gazing in AI research at the expense of examining current technology is explored by Crawford and Calo [29].

TABLE I
TAXONOMY OF METHODS TO CONSTRAIN A-BOT BEHAVIOUR

AI Type	**Strengths**	**Weaknesses**	**Associated Methods**	
			Explicit Rules	**Implicit Rules (Requires Data)**
Symbolic	Efficient generalisation. Interpretable	Training difficult	Model Verification, Program Verification, CP-nets	Sequential association rule mining?
Statistical (Soft)	Training easy, Software freely available	Data intensive. Poor generalisation. Interpretation difficult	Safe RL, Seldonian Learning	IRL, RL

policy found through unconstrained RL. Abel et al [38] present a related RL method to infer the reward function of an ethical demonstrator and create an ideal ethical agent. Training sidesteps the requirement for an explicit representation of rules; A-bots instead learn them implicitly. The weakness of this approach when used alone is that the expert behaviour observed, or the simulated experience gathered in training, or experience gathered when deployed must comprehensively demonstrate all laws. In the related and more popular subject area of ethically constraining A-bots, Rossi and Mattei [39], call these approaches data driven in contrast with symbolic rule based approaches which we will now briefly discuss.

Reasoning is the approach where the A-bot has a rule-set and the ability to reason about the legality of its actions. Alves et al [6] successfully demonstrate a way of encoding UK road junction rules using temporal logic and then test the behaviour of an A-bot implemented as a Belief-Desire-Intention (BDI) Agent and testing its behaviour exhaustively against the ruleset using program verification techniques adapted for Agent based programming. Rossi and Mattei also place CP-nets in this category of symbolic approach. It is a method of ordering preferences over possible actions.

The oracle structure presented in this paper is a mix of Symbolic and Soft AI. Noothigattu et al [37] constrain behaviour via a legal oracle which informs the A-bot of the legality of an action through a binary signal. This is done with contextual bandits so the policies and rules considered are limited in complexity to single period rules and policies. The aforementioned RL-Shield papers of [20] and [21] are examples of this mixed approach. Techniques to learn through RL subject to rules expressed in temporal symbolic logic are also presented in [40], [41] and [42].

VI. CONCLUSION

This paper introduces the experimental concept of an AI powered Legal counsel for Autonomous Algorithmic agents (A-bots). This structure can constrain the behaviour of an autonomous algorithmic agent or A-bot to be legal through the presence of a Legal Oracle (and related Rule book). Its use is dual purpose. Firstly in training it can impart an explicit rule-set to an A-bot of relatively flexible design. Secondly it can be used on deployment as a contingency system to cover situations which the A-bot may not have met in training. I have identified five capabilities of the Legal Oracle which are necessary for it to operate in a regulated environment with complex laws including laws of specific intent and inchoate offences. Shortly, I aim to demonstrate a working version of the Legal Oracle to train an A-bot to learn a specific regulated function.

REFERENCES

[1] M. P. Wellman and U. Rajan, "Ethical Issues for Autonomous Trading Agents," *Minds and Machines*, vol. 27, no. 4, pp. 609–624, 2017.

[2] V. Manahov, "Front-Running Scalping Strategies and Market Manipulation: Why Does High-Frequency Trading Need Stricter Regulation?" *Financial Review*, vol. 51, no. 3, pp. 363–402, 2016.

[3] E. Calvano, G. Calzolari, V. Denicolo, and S. Pastorello, "Artificial Intelligence, Algorithmic Pricing and Collusion," 2019. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3304991

[4] A. Etzioni and O. Etzioni, "Viewpoint designing ai systems that obey our laws and values calling for research on automatic oversight for artificial intelligence systems." *Communications of the ACM*, vol. 59, no. 9, pp. 29–31, 2016.

[5] H. Prakken, "On how AI & law can help autonomous systems obey the law : a position paper," *AI4J-Artificial Intelligence for Justice*, pp. 42–46, 2016.

[6] G. V. Alves, L. Dennis, and M. Fisher, "Formalisation and Implementation of Road Junction Rules on an Autonomous Vehicle Modelled as an Agent," in *Formal Methods. FM 2019 International Workshops*, vol. 1. Springer International Publishing, 2020, pp. 217–232.

[7] G. Boella, L. Humphrey, R. Muthuri, P. Rossi, and L. Van Der Torre, "A critical analysis of legal requirements engineering from the perspective of legal practice," *2014 IEEE 7th International Workshop on Requirements Engineering and Law, RELAW 2014 - Proceedings*, pp. 14–21, 2014.

[8] IEEE, "Embedding Values into Autonomous Intelligent Systems," pp. 33–54, 2016.

[9] J. Turner, *Robot Rules*. Palgrave Macmillan, 2019.

[10] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, ..., and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[11] J. Y. Halpern, *Actual Causality*. MIT Press, 2016.

[12] X. Wang and M. P. Wellman, "Spoofing the Limit Order Book: An Agent-Based Model," *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems. International Foundation for Autonomous Agents and Multiagent Systems*, pp. 651–659, 2017. [Online]. Available: http://strategicreasoning.org/wp-content/uploads/2017/03/spoofing-limit-order.pdf

[13] Y. Bathaee, "The artificial intelligence black box and the failure of intent and causation," *Harvard Journal of Law & Technology*, vol. 2, no. 4, pp. 31–40, 2011. [Online]. Available: https://www.theverge.com/

[14] S. Chopra and L. F. White, *A Legal Theory for Autonomous Artificial Agents*. University of Michigan Press, 2016. [Online]. Available: https://doi.org/10.3998/mpub.356801

[15] H. Ashton, "Definitions of intent for AI derived from common law," in *Jurisin 2020: 14th International Workshop on Jurisi-informatics*, 2020. [Online]. Available: https://easychair.org/publications/preprint/GfCZ

[16] H. Prakken, "On the problem of making autonomous vehicles conform to traffic law," *Artificial Intelligence and Law*, vol. 25, no. 3, pp. 341–363, 2017.

[17] M. Hasanbeig, A. Abate, and D. Kroening, "Certified reinforcement learning with logic guidance," 2019. [Online]. Available: https://arxiv.org/pdf/1902.00778.pdf

[18] A. F. T. Winfield, C. Blum, and W. Liu, "Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection," in *Advances in Autonomous Robotics Systems.*, 2014, pp. 85–96.

[19] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, ..., and D. Hassabis, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017. [Online]. Available: http://dx.doi.org/10.1038/nature24270

[20] N. Jansen, S. Junges, K. Bettina, and R. Bloem, "Shielded Decision-Making in MDPs," 2018.

[21] M. Alshiekh, R. Bloem, K. Bettina, S. Niekum, U. Topcu, and E. Street, "Safe Reinforcement Learning via Shielding," in *AAAI Conference on Artifical Intelligence*, 2018. [Online]. Available: https://aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17211/16534

[22] C. Baier and J.-P. Katoen, *Principles Of Model Checking*. MIT Press, 2008. [Online]. Available: http://mitpress.mit.edu/books/principles-model-checking

[23] V. Charisi, L. Dennis, M. Fisher, R. Lieck, A. Matthias, ..., and R. Yampolskiy, "Towards Moral Autonomous Systems," pp. 1–34, 2017. [Online]. Available: http://arxiv.org/abs/1703.04741

[24] T. Arnold, D. Kasenberg, and M. Scheutz, "Value alignment or misalignment - What will keep systems accountable?" in *AAAI Workshop - Technical Report*, 2017. [Online]. Available: https://hrilab.tufts.edu/publications/aaai17-alignment.pdf

[25] A. F. T. Winfield, K. Michael, J. Pitt, and V. Evers, "Machine ethics: The design and governance of ethical ai and autonomous systems," *Proceedings of the IEEE*, vol. 107, no. 3, pp. 509–517, 2019.

[26] R. Leenes and F. Lucivero, "Laws on robots, laws by robots, laws in robots: Regulating robot behaviour by design," *Law, Innovation and Technology*, vol. 6, no. 2, pp. 193–220, 2014.

[27] P. Lin, "Why Ethics Matters for Autonomous Cars," in *Autonomous Driving: Technical, Legal and Social Aspects*. Springer Open, 2016, ch. 5, pp. 69–85.

[28] A. Theodorou and V. Dignum, "Towards ethical and socio-legal governance in AI," *Nature Machine Intelligence*, vol. 2, pp. 10–12, 2020.

[29] K. Crawford and R. Calo, "There is a blind spot in AI research," *Nature*, vol. 538, no. 7625, pp. 311–313, 2016.

[30] M. Hildebrandt, "Closure: on ethics, code and law," in *Law for Computer Scientists*. Oxford University Press, 2019, ch. 11.

[31] C. Cath, "Governing artificial intelligence: Ethical, legal and technical opportunities and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, 2018.

[32] W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, 2009.

[33] J. García and F. Fernández, "A Comprehensive Survey on Safe Reinforcement Learning," *Journal of Machine Learning Research*, vol. 16, pp. 1437–1480, 2015.

[34] P. S. Thomas, B. C. da Silva, A. G. Barto, S. Giguere, Y. Brun, and E. Brunskill, "Preventing undesirable behavior of intelligent machines," *Science*, vol. 366, no. 6468, 2019.

[35] J. Lehman, J. Clune, and D. Misevic, "The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities," *Artificial Life*, vol. 26, no. 2, pp. 274–306, 2020.

[36] P. Abbeel and A. Y. Ng, "Apprenticeship Learning via Inverse Reinforcement Learning," in *Proceedings of the 2st International Conference on Machine Learning (ICML)*, 2004. [Online]. Available: https://ai.stanford.edu/ ang/papers/icml04-apprentice.pdf

[37] R. Noothigattu, D. Bouneffouf, N. Mattei, R. Chandra, P. Madan, ..., and F. Rossi, "Interpretable Multi-Objective Reinforcement Learning through Policy Orchestration," 2018. [Online]. Available: http://arxiv.org/abs/1809.08343

[38] D. Abel, J. MacGlashan, and M. L. Littman, "Reinforcement learning as a framework for ethical decision making," *AAAI Workshop - Technical Report*, vol. WS-16-01 -, pp. 54–61, 2016.

[39] F. Rossi and N. Mattei, "Building Ethically Bounded AI," in *Proceeding of the AAAI Conference on Artificial Intelligence*, 2019, pp. 9785–9789. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.33019785

[40] G. Mason, R. Calinescu, D. Kudenko, and A. Banks, "Assured reinforcement learning with formally verified abstract policies," in *ICAART 2017 - Proceedings of the 9th International Conference on Agents and Artificial Intelligence*, vol. 2, 2017, pp. 105–117.

[41] M. Hasanbeig, A. Abate, and D. Kroening, "Logically-Constrained Reinforcement Learning," 2019. [Online]. Available: https://arxiv.org/abs/1801.08099

[42] M. Wen, R. Ehlers, and U. Topcu, "Correct-by-synthesis reinforcement learning with temporal logic constraints," *IEEE International Conference on Intelligent Robots and Systems*, pp. 4983–4990, 2015.